



UNIVERSIDADE DA CORUÑA

Facultade de Economía e Empresa

Traballo de
fin de grao

Análise da metodoloxía
estatística empregada
na Enquisa de
Poboación Activa

Leticia Lema López

Titor: Xosé Manuel Martínez
Filgueira

Grao en Economía

Ano 2013

Traballo de Fin de Grao presentado na Facultade de Economía e Empresa da Universidade da Coruña
para a obtención do Grao en Economía

Resumo

Dado que a Economía é unha ciencia social que depende en gran medida de datos, a Estatística convértese nunha ferramenta de primeira necesidade. Temos que acudir a ela ante a utilidade de dispoñer dun indicador que mida os compoñentes referentes ó mercado de traballo da maneira máis eficaz posible.

A Enquisa de Poboación Activa (EPA) proporciónanos información moi veraz do mercado laboral. Realízase baixo os supostos da Teoría de Mostras, a cal nos permite coñecer a realidade entrevistando só a unha parte da poboación que sexa representativa do total. En España, aplícase unha mostraxe estratificada en dúas etapas. Consiste en dividir a poboación en áreas xeográficas (unidades da primeira etapa), en función de criterios demográficos e socioeconómicos, facendo a mostraxe de cada área xeográfica aleatoriamente, para obter a parte proporcional da mostra. A segunda etapa consiste na división das áreas xeográficas por vivendas, e a mostraxe aleatoria destas últimas. Deste modo, logramos obter resultados medibles de gran calidade; isto é así grazas a que a calidade de ditos resultados é medida a través dos erros de mostraxe, os cales son conseguidos mediante a varianza do estimador.

No presente traballo analízase o método estatístico aplicado na Enquisa de Poboación Activa en distintos países, especialmente nos da Unión Europea. Analízase con máis detalle a Enquisa de Poboación Activa en España, expoñendo todo o relativo o tipo de mostraxe aplicado, os aspectos necesarios para a súa aplicación (ámbito da enquisa, definicións, taxas,...) e os constantes cambios realizados na enquisa coa finalidade de seguir aumentando a súa calidade. Préstase especial atención a obtención dos erros de mostraxe, realizando unha aplicación práctica onde se obteñen.

Palabras clave: enquisa de poboación activa, teoría de mostras, mercado de traballo, erro de mostraxe, estatística laboral.

Número de palabras: 14.018

Abstract

Given that the Economy is a social science that depends largely on data, the Statistic becomes a tool of first necessity. We have to resort to it facing the utility of to have an indicator to measure the components relating to the labor market the most effective way possible.

The Labour Force Survey (LFS) provides very veracious information of the labor market. Is performed under the supposed of the Theory of Samples, which allows us to know the reality by interviewing only one part of the population that is representative of the total. In Spain, stratified sampling is applied in two stages. It consists of to divide the population into geographic areas (units of the first stage), according to demographic and socioeconomic criteria, making the sampling of each geographical area at random, to get the proportion of the sample. The second stage consists in the division of geographical areas by housing, and the random sampling of the latter. Thus, we managed to get measurable results of high quality, so this is thanks to the quality of these results is measured through sampling errors, which are achieved by the variance of the estimator.

This work analyzes the statistical method applied to the Labour Force Survey in different countries, especially in the European Union. Is analyzed in more detail the Labour Force Survey in Spain, exposing all regarding the type of sampling applied, aspects required for its application (scope of the survey, definitions, rates, ...) and the constant changes in the survey with the aim of to continue increasing its quality. Special attention is given to obtain the sampling errors, performing a practical application where the sampling errors are obtained.

Índice

Introdución.....	5
1. Teoría de Mostras	7
2. Estatística laboral: Enquisa de Poboación Activa.....	19
2.1. Introducción á Enquisa de Poboación Activa.....	19
2.2. A Enquisa de Poboación activa en diferentes países	21
2.3. A Enquisa de Poboación activa en España.....	31
3. Medición dos erros de mostraxe: aplicación práctica.....	45
Conclusións	50
Anexos.....	53
Bibliografía.....	65

Introdución

A Economía é unha ciencia social xa que se dedica ao estudo dos procedementos produtivos e de intercambio e á análise do consumo de bens e servizos que realizan as persoas. En 1932, o británico Lionel Robbins aportou unha definición de ciencia económica ao considerala como a rama da ciencia que analiza como os seres humanos compracen as súas necesidades ilimitadas con recursos escasos que teñen diferentes usos (Steven, G. Medema e Warren, J. Samuels, 1988).

Para levar a cabo a súa función de aportar criterios racionais para a asignación de recursos escasos, a Economía necesita traballar con datos e de aí o interese que para ela ten a Estatística. Esta última é unha rama da ciencia encargada do deseño de experimentos ou procedementos de mostraxe, da análise de datos e dos procedementos para inferir acerca dunha poboación de medicións con base na información contida na mostra. A estatística ofrece unha gran variedade de ferramentas analíticas na toma de decisións baixo incertidumbre (Mendenhall e Reinmuth, 1978). Segundo nos di Peña, D. (2001), esta ciencia actualmente é o resultado da unión de dúas disciplinas que evolucionan independentemente ata confluir no s. XIX: a primeira é o cálculo de probabilidades e a segunda é a “estatística” que estuda a descripción de datos. A integración de ambas liñas de pensamento da lugar a unha ciencia que estuda cómo obter conclusións da investigación empírica mediante o uso de modelos matemáticos.

Etimolóxicamente a palabra Estatística ten a súa raíz na palabra “estadista”, ésta á súa vez no latín “status” e refírese o interese de coñecer a situación do “estado”

(Martín Pliego, F.J., 1994). Herdeira dese concepto é a estatística oficial, que por medio de oficinas públicas proporciona aos cidadáns, ás autoridades públicas e ós axentes económicos e sociais, información de calidade e fiable que nos permite explicar a realidade económica e a súa evolución no tempo.

Unha área estatística de vital importancia para a recollida de información é a Teoría de Mostras. Cando o investigador toma información de todos e cada un dos elementos da poboación dicimos que está realizando un censo. Pero isto non é posible en moitas ocasións, por exemplo polo custo que supoñería, polo que o investigador toma información a uns cantos elementos da poboación estatística, chamándose este proceso mostraxe. Segundo Muruzábal, J.J. (2005), a Teoría de Mostras xunto coa Inferencia Estatística ocúpanse da análise mostral dos fenómenos aleatorios en tres vertentes: no deseño das mostras, para garantir que estas son representativas e presentan a dimensión axeitada en razón dos obxectivos da investigación; na recompilación e presentación dos datos obtidos, permitindo un coñecemento detallado da parte do universo que foi estudado e no análise mostral e a extensión de conclusións ó conxunto do fenómeno, para alcanzar un coñecemento do mesmo a partir dos valores mostrais mediante un razoamento indutivo con fundamento matemático.

Sendo conscientes da relevancia que un bo funcionamento do mercado laboral ten para a Economía, analizarase neste traballo o seu principal indicador estatístico: a Enquisa de Poboación Activa (a partir de agora, EPA), facendo especial fincapé nos aspectos da Teoría de Mostras aplicados na súa elaboración. Para iso, previamente introducirase á Teoría de Mostras, cos distintos tipos de mostraxe que se poden levar a cabo. Comezando polos seus aspectos teóricos e continuando coa descrición da súa aplicación en distintos países, tanto da zona euro como non europeos, na que se compararán diferentes aspectos da súa elaboración, aínda que se dará máis relevancia ao seu deseño mostral. Rematarase esta parte cunha descrición máis minuciosa da elaboración da EPA en España.

No derradeiro capítulo complementarase a descrición anterior, prestando especial atención aos aspectos da calidade, cunha aplicación práctica, na cal se empregarán os conceptos descritos para a EPA en España, para amosar a metodoloxía de cálculo do erro de mostraxe. Finalmente, expoñeranse unhas conclusións e suxestións persoais sobre todo o estudo que sintetizen todo o traballo realizado.

1. Teoría de Mostras

Para chegar ao noso último fin de obter resultados que nos permitan decidir, temos que comezar recollendo datos. Á hora de facelo temos dúas posibilidades: a primeira, facer unha análise exhaustiva analizando toda a poboación (o que chamamos censo); a segunda sería usar unha mostra. Esta última opción permítenos coñecer a realidade entrevistando a unha parte da poboación; xa que nalgúns casos analizar toda a poboación pode ser imposible debido os custos que supón e tamén ó tempo e esforzo que require.

Debido a que a calidade da mostra recollida influirá sobre a calidade das conclusións obtidas, a selección dos individuos da mostra deberá ser imparcial e tecnicamente boa. A mostra terá que ser bastante ampla como para permitir coñecer a realidade, pero tampouco ten que ser forzosamente grande, é dicir, o seu tamaño ten que ser suficiente.

Os modelos estatísticos dependen de certas constantes descoñecidas que chamaremos parámetros (Peña, D. 2001). O obxectivo da mostraxe é facer inferencia sobre eses parámetros que pertencen a unha poboación.

No que segue imos introducir os conceptos básicos que manexa a Teoría de Mostras e posteriormente os seus principais procedementos o que nos facilitará máis tarde a análise da súa aplicación para a EPA.

A partir da información exposta por: Casal, J. (2003) e Alba Fernández, V. e Ruiz Fuentes, N. (2004) e Sharon L. Lohr (1999), derívanse os seguintes conceptos:

- A poboación total ou poboación obxectivo é o grupo de individuos do que se pretende obter información (no caso que nos interesa serían todos os individuos de 16 a 74 anos que forman o mercado de traballo).
- A poboación investigada: xa que a miúdo a poboación non é accesible na súa totalidade debemos traballar só sobre unha parte dela, que será a poboación estudada, da cal obtemos a mostra. Na EPA, serán todos os individuos que finalmente son investigados e polo tanto, forman parte da mostra.
- O marco da enquisa: é o listado dos individuos da poboación. (as veces non vai ser necesario dispoñer de todo o listado).
- Unidade da enquisa: é cada individuo da poboación estudada. Segundo o tipo de mostraxe poderemos diferenciar entre unidades primarias, secundarias, etc. Será a persoa ou o grupo de vivendas que finalmente é enquisado.
- Fracción da enquisa: é unha proporción de individuos da poboación estudada que forma parte da mostra.
- Sesgo: son os erros sistemáticos que se producen (diferentes dos erros de estimación).

Outro concepto importante para manexar na Teoría de Mostras é o do erro.

Cando temos uns resultados, necesitamos saber se estes son de calidade, é dicir, se son fiables, para iso o “erro” mídenos o que se aparta a estimación do valor real, o cal é descoñecido. Os erros que afectan a enquisa agruparémolos en dous grupos, seguindo a clasificación e o exposto polo (Instituto Nacional de Estadística [INE], 2012):

- Erros alleos á mostraxe: (son comúns a toda investigación estatística, tanto se a información é recollida por mostraxe como se se realiza un Censo). Estes erros preséntanse en calquera fase do proceso estatístico: antes da toma de datos (por exemplo, por deficiencias do marco e insuficiencias nas definicións e cuestionarios), durante a toma de datos (por deficiencias na

labor dos entrevistadores e unha incorrecta declaración por parte dos entrevistados) e tras a recollida de datos (erros na depuración, codificación, gravación ou tabulación dos resultados). Este erro pódese reducir de forma importante coidando os detalles, crece co tamaño da mostra, pero si se pode reducir planificando ben a selección da mostra: dando instrucións precisas e claras, proporcionando cuestionarios redactados con precisión e fáciles de responder ou seleccionando a persoal moi cualificado.

- Erros de mostraxe: orixínanse pola obtención de resultados sobre as características da poboación, a partir da información recollida nunha mostra dela. Este erro ten a vantaxe de ser cuantificable e poder ser analizado con ferramentas matemáticas. Este tipo de erro ten as características de que non se pode eliminar, depende do tamaño da mostra decrecendo se a mostra se amplía e é menor canto máis homoxénea sexa a poboación, ou o que é o mesmo, canto menos homoxéneos sexan os datos máis disperso será o resultado obtido, polo tanto podemos facer a medición deste erro mediante medidas de dispersión:

Erro de mostraxe absoluto: $\hat{E} = \hat{\sigma} \hat{\theta}$ (para un estimador $\hat{\theta}$)

Erro de mostraxe relativo: $\hat{e} = \frac{\hat{E}}{\theta}$ (sóese dar en %)

A medida que se emprega para coñecer o erro, en xeral, soe ser a desviación típica do estimador (ou este mesmo ó cadrado que é a varianza), e o seu equivalente en termos relativos é o coeficiente de variación de Pearson.

A Teoría de Mostras permítenos coñecer o nivel de erro cometido en situacións controladas, iso permítenos construír un intervalo de confianza para indicar o nivel de calidade dunha enquisa.

Segundo nos di Rodríguez, R.J. (1996), a inferencia estatística é o proceso consistente en facer uso dos resultados mostrais para obter conclusións sobre as características dunha poboación. Empréganse procedementos estatísticos que permitan estimar a proporción (%) real dunha poboación, é dicir, utilízanse estatísticos como estimadores dos parámetros (poboacionais) mediante un razoamento inductivo

ou de xeneralización estatística. Dita inferencia realízase a través dun procedemento de estimación por intervalo.

O intervalo de confianza é un conxunto de resultados nos que confiamos que conteñan (con certa marxe de probabilidade) os valores verdadeiros da poboación. Esta marxe de probabilidade mídese mediante o coeficiente de confianza (β), que é a medición probabilística de que o intervalo fixado polo erro de mostraxe conteña o parámetro que se desexa estimar.

O nivel de significación (representado pola letra grega α) é o complemento á unidade do coeficiente de confianza ($\beta + \alpha = 100\%$). Para un coeficiente de confianza dado conseguimos un valor z , o cal, en combinación cunha medida de variabilidade da mostra vainos dar a amplitude do intervalo de confianza, e polo tanto permitíranos construílo. Así mesmo, ese valor z intervén tamén no cálculo dos tamaños mostrais.

O que se obtén mediante a estimación por intervalo é unha expresión do tipo:

$$\text{Parámetro mostral } (P_m) \pm \text{erro de estimación (e \%)}$$

O intervalo de confianza viría dado pola seguinte expresión:

$$P(\hat{p} - z\sqrt{\widehat{\text{var}}(\hat{p})} < p < \hat{p} + z\sqrt{\widehat{\text{var}}(\hat{p})}) = \beta$$

Existen diferentes tipos de mostraxe que serán útiles dependendo da poboación da que queiramos sacar conclusións. Verémolo, no noso caso, cando se faga a comparación con outros países na metodoloxía da EPA, e dependendo do país, o tipo de mostraxe empregado será diferente.

Pérez, C. (1999), defínenos os métodos de mostraxe como o conxunto de técnicas estatísticas que estudan a forma de seleccionar unha mostra o suficientemente representativa dunha poboación, cuxa información permita inferir as propiedades ou características de toda a poboación cometendo un erro medible e acotable.

Os métodos de mostraxe poden dividirse en dous grandes grupos: probabilísticos e non probabilísticos. Os probabilísticos son aqueles nos cales a selección é aleatoria, coñecendo as probabilidades de que cada elemento sexa seleccionado. Entre as ventaxas destes métodos está a de permitir coñecer o erro estándar que se comete ao facer as estimacións, desta forma pódese dar unha medida da representatividade da

mostra sobre o total da poboación. Pola contra, os non probabilísticos son aqueles métodos nos que non se realiza unha selección aleatoria, polo que non se pode garantir propiedades probabilísticas.

Segundo a información recollida dos autores, Casal, J. (2003), Peña, D. (2001) e Pérez, C. (1999), existen os seguintes tipos de mostraxe non probabilística:

- Mostraxe por selección intencionada ou mostraxe por conveniencia: Consiste na elección por métodos non aleatorios, dunha mostra cuxas características sexan similares ás da poboación obxectivo. A representatividade é determinada polo investigador de modo subxectivo, sendo este o maior inconveniente deste método. Presenta case sempre sesgos polo tanto non é eficiente empregalo. Pode ser útil cando se pretende realizar unha primeira prospección da poboación ou cando non existe un marco da enquisa definido. Neste tipo de mostraxe nunca se representará a variabilidade da poboación, que normalmente quedará subestimada.
- Mostraxe de xuízos (ou opinático, ou a criterios): Consiste en que o investigador selecciona a aqueles individuos que considera expertos na materia de interese. Estes expertos van ter opinións diferentes e non hai ningún método obxectivo para diferenciar entre as mesmas nin para medir o seu grao de erro. A persoa que selecciona a mostra é a que procura que sexa representativa, dependendo esta última da súa opinión, polo que a avaliación da representatividade é subxectiva.
- Mostraxe por cotas: Clasifícase a poboación segundo diferentes criterios, (sexo, idade, nivel de estudos, nivel económico...), do mesmo xeito que se fai coa mostraxe estratificada, sen embargo, á hora de seleccionar os individuos para cada grupo non se seguen criterios aleatorios, se non máis ben de comodidade.
- Mostraxe de “bola de neve”: Usado con poboacións especiais, ou de difícil acceso. A partir dunha selección previa de elementos, súxíreselle a estes que indiquen outros membros do grupo cos que continuar a investigación, o mesmo tempo que se lles pode pedir que actúen de intermediarios.

Este tipo de mostraxes non probabilísticas soen aplicarse a miúdo na vida corrente (sobre todo no comercio e en enquisas de opinión), e sempre que en caso de

equivocación as consecuencias non sexan demasiado graves. Por isto, no ámbito que nos concerne, do mercado laboral, soen aplicarse métodos probabilísticos, xa que tamén é fundamental que o método de mostraxe aplicado nos deixe analizar a calidade da enquisa, o cal só acontece cos métodos probabilísticos.

Segundo a información recollida dos autores, Casal, J. (2003), Peña, D. (2001) e Pérez, C. (1999), existen os seguintes tipos de mostraxe probabilística:

- Mostraxe aleatoria: Todos os elementos teñen a mesma probabilidade de ser elixidos. Os individuos que forman a mostra son elixidos ó azar mediante números aleatorios (utilizando táboas de números aleatorios ou xerándoos mediante ordenador). Pode realizarse de distintas formas, sendo as máis frecuentes a mostraxe simple, sistemática, estratificada e por conglomerados.
- Mostraxe aleatoria simple: É o método conceptualmente máis simple. Consiste en extraer todos os individuos ó azar do marco da enquisa, é dicir, con probabilidades iguais de ser seleccionados. Na práctica, a non ser que se trate de poboacións pequenas ou de estrutura moi simple, é difícil de levar a cabo eficientemente.

Esta mostraxe pode ser sen reempazamento ou con reempazamento. Na primeira delas, o procedemento de selección con probabilidades iguais consiste na obtencion da mostra unidade a unidade de forma aleatoria sen reempazamento á poboación das unidades previamente seleccionadas; ademáis a orde de colocación dos elementos nas mostras non intervéñen (mostras cos mesmos elementos colocados en distinta orde considéranse iguais).

Este procedemento de mostraxe é o máis simple conceptualmente polo cal, aínda que non sexa moi frecuente o seu uso na práctica, si é amplamente usado como referencia para os fundamentos teóricos da Teoría de Mostras (Levy, P. S., e Lemeshow, S., 1999). Ese é o motivo de que se presenten aquí as expresións para a estimación da media e a proporción, así como as varianzas mostrais e os intervalos de confianza, algo que no resto dos procedementos só se repetirá na mostraxe estratificada por ser a que se emprega no INE para recoller a mostra da EPA.

Na mostraxe aleatoria simple sen reempazamento, sendo π_i a probabilidade de que a unidade u_i pertenza a mostra, temos que $\pi_i = n/N$, polo que os estimadores lineais insesgados para os parámetros máis comúns serán:

$$\theta = X = \sum_{i=1}^N X_i \rightarrow \hat{\theta} = \hat{X} = \sum_{i=1}^n \frac{X_i}{\pi_i} = \sum_{i=1}^n \frac{X_i}{\frac{n}{N}} = N \frac{1}{n} \sum_{i=1}^n X_i = N\bar{x}$$

$$\theta = \bar{X} = \sum_{i=1}^N \frac{X_i}{N} \rightarrow Y_i = \frac{X_i}{N} \rightarrow \hat{\theta} = \hat{\bar{X}} = \sum_{i=1}^n \frac{\frac{X_i}{N}}{\pi_i} = \sum_{i=1}^n \frac{\frac{X_i}{N}}{\frac{n}{N}} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{x}$$

$$\theta = P = \sum_{i=1}^N \frac{A_i}{N} \rightarrow \hat{\theta} = \hat{P} = \sum_{i=1}^n \frac{\frac{A_i}{N}}{\pi_i} = \frac{1}{n} \sum_{i=1}^n A_i = \hat{P}$$

Baseándonos en que a varianza é invariante ante un cambio na orixe:

$$V(\hat{X}) = V(\hat{X} - \bar{X})$$

Sendo $f = n/N$, chegamos á fórmula clásica para a varianza do estimador:

$$V(\hat{X}) = N^2(1-f) \frac{S^2}{n}$$

O intervalo de confianza, na estimación de X , nesta mostraxe será:

$$P\left(\hat{X} - z \sqrt{N^2(1-f) \frac{S^2}{n}(\hat{X})} < X < \hat{X} + z \sqrt{N^2(1-f) \frac{S^2}{n}(\hat{X})}\right) = \beta$$

Pola contra, a mostraxe aleatoria con reempazamento consiste nun procedemento de selección con probabilidades iguais consistente en obter da mostra unidade a unidade de forma aleatoria con reempazamento á poboación das unidades previamente seleccionadas. As mostras con elementos repetidos son posibles e calquer elemento da poboación pode estar repetido 0,1..., n veces.

Repóñense á poboación as unidades previamente seleccionadas e con probabilidades iguais, realízase a selección sucesiva das unidades para a mostra con probabilidades $P_i = 1/N$ e todas as mostras son equiprobables.

Dado que na mostraxe aleatoria simple con reempazamento temos que $P_i = 1/N$, os estimadores lineais insesgados para os parámetros poboacionais máis comúns serán:

$$\theta = X = \sum_{i=1}^N X_i \rightarrow \hat{\theta} = \hat{X} = \sum_{i=1}^n \frac{X_i}{nP_i} = \sum_{i=1}^n \frac{X_i}{\frac{n}{N}} = N \frac{1}{n} \sum_{i=1}^n X_i = N\bar{X}$$

$$\theta = \bar{X} = \sum_{i=1}^N \frac{X_i}{N} \rightarrow Y_i = \frac{X_i}{N} \rightarrow \hat{\theta} = \hat{Y} = \sum_{i=1}^n \frac{Y_i}{nP_i} = \sum_{i=1}^n \frac{Y_i}{\frac{n}{N}} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$$

$$\theta = P = \sum_{i=1}^N \frac{A_i}{N} \rightarrow \hat{\theta} = \hat{P} = \sum_{i=1}^n \frac{A_i}{\frac{n}{N}} = \frac{1}{n} \sum_{i=1}^n A_i = \hat{P}$$

A varianza do estimador será:

$$V(\hat{X}) = N^2 \frac{\hat{S}^2}{n}$$

O intervalo de confianza para o parámetro X será o seguinte:

$$P\left(\hat{X} - z \sqrt{N^2 \frac{\hat{S}^2}{n}} < X < \hat{X} + z \sqrt{N^2 \frac{\hat{S}^2}{n}}\right) = \beta$$

- Mostraxe sistemática: Elíxese o primeiro individuo ó azar e o resto vén condicionado por él. Consideramos unha poboación de tamaño N, e agrupamos os seus elementos en n zonas de tamaño k ($N=nk$), para extraer unha mostra de tamaño n, elíxese ó azar unha unidade na primeira zona, e

para seleccionar as $n-1$ unidades restantes para a mostra, tómasse en cada zona a unidade que ocupa o mesmo lugar dentro da súa zona que o que ocupaba a primeira unidade seleccionada dentro da primeira zona

Este método é moi simple de aplicar na práctica e ten a ventaxa de non ter que dispoñer dun marco da enquisa elaborado. Pode aplicarse na maioría das situacións, a única precaución que debe terse en conta é comprobar que a característica que estudamos non teña unha periodicidade que coincida coa da mostraxe, xa que nese caso aumentaría a varianza. Outro inconveniente é que, en xeral, só vai existir selección aleatoria para a primeira unidade da mostra.

- Mostraxe aleatoria estratificada: Divídese a poboación en grupos en función dun carácter determinado e despois mostrease a cada grupo aleatoriamente, para obter a parte proporcional da mostra. Se sabemos con seguridade que un determinado estrato aporta unidades para a mostra, dito estrato denomínase estrato correpresentado. As unidades da poboación que con certeza van a pertencer á mostra denomínanse unidades autorrepresentadas.

Este método aplícase para evitar que por azar algún grupo de individuos estea menos representado cós outros. Ten interés cando a característica en cuestión pode estar relacionada coa variable que queremos estudar. Pode aportar información máis precisa de algunhas subpoboacións que varían bastante en tamaño e propiedades entre si, pero que son homoxéneas dentro de si.

Ademais pode permitir ganancias en precisión xa que estratos máis homoxéneos reducen a variabilidade. Os estratos deberían no posible estar constituídos por unidades homoxéneas, xa que no caso límite de estricta homoxeneidade bastaría con seleccionar unha soa unidade en cada estrato.

Neste método, a afixación é o reparto, asignación ou distribución do tamaño mostral entre os diferentes estratos. Pódense dar os seguintes tipos de afixación:

A afixación uniforme consiste en asignar o mesmo número de unidades mostrais a cada estrato, polo que todos os estratos terán o mesmo tamaño (o cal vai favorecer os estratos de menor tamaño e perxudicar aos grandes en canto a precisión, polo cal é conveniente en poboacións con estratos do mesmo tamaño).

A afixación proporcional consiste en asignar a cada estrato un número de unidades mostrais proporcional ó seu tamaño. As fraccións de mostraxe nos estratos van ser iguais, coincidindo coa fracción global de mostraxe.

A afixación de mínima varianza consiste en determinar os valores do número de unidades que se extraen dun determinado estrato para a mostra, de modo que para un tamaño de mostra fixo, a varianza dos estimadores sexa mínima.

Por último, a afixación óptima consiste en determinar os valores da mostra de forma que para un coste fixo, a varianza dos estimadores sexa mínima.

En primeiro lugar, considerando que seleccionamos en cada estrato as unidades para a mostra mediante mostraxe aleatoria simple sen reempazamento (polo que a selección realizarase de forma independente en cada estrato).

Donde π_{hi} é a probabilidade de que a unidade u_{hi} pertenza a mostra \tilde{X}_h , de n_h unidades, obtidas de entre as N_h unidades do estrato h-ésimo.

Tendo presente que:

$$\pi_{hi} = \frac{n_h}{N_h} = f_h$$

A estimación do total poboacional será:

$$\theta = X \rightarrow Y_{hi} = X_{hi} \rightarrow \hat{X}_{st} = \sum_{h=1}^L \sum_{i=1}^{n_h} \frac{X_{hi}}{\pi_{hi}} = \sum_{h=1}^L \frac{1}{N} \sum_{i=1}^{n_h} \frac{X_{hi}}{n_h/N_h} = \sum_{h=1}^L \hat{X}_h$$

O estimador da media poboacional na mostraxe estratificada aleatoria é a media ponderada dos estimadores da media en cada estrato, sendo os coeficientes de ponderación: $W_h = N_h/N$.

$$\theta = \bar{X} \rightarrow Y_{hi} = \frac{X_{hi}}{N} \rightarrow \bar{X}_{st} = \sum_{h=1}^L \frac{1}{N} \sum_{i=1}^{n_h} \frac{X_{hi}}{\pi_{hi}} = \sum_{h=1}^L \frac{1}{N} \sum_{i=1}^{n_h} \frac{X_{hi}}{n_h/N_h} = \sum_{h=1}^L \frac{N_h}{N} \frac{1}{n_h} \sum_{i=1}^{n_h} X_{hi} = \sum_{h=1}^L W_h \bar{x}_h$$

A estimación da proporción:

$$\theta = P \rightarrow Y_{hi} = \frac{A_{hi}}{N} \rightarrow \bar{P}_{st} = \sum_{h=1}^L \frac{1}{N} \sum_{i=1}^{n_h} \frac{A_{hi}}{\pi_{hi}} = \sum_{h=1}^L \frac{1}{N} \sum_{i=1}^{n_h} \frac{A_{hi}}{n_h/N_h} = \sum_{h=1}^L w_h \bar{P}_h$$

A varianza do estimador é igual a suma das varianzas das estimacións dos totais de cada estrato, xa que a mostraxe sen reempazamento, realízase de forma independente nos distintos estratos.

$$V(\hat{X}_{st}) = V\left(\sum_{h=1}^L \hat{X}_h\right) = \sum_{h=1}^L N_h^2 \cdot (1 - f_h) \cdot \frac{S_h^2}{n_h}$$

O intervalo de confianza para o parámetro X será o seguinte:

$$P\left(\hat{X} - z \sqrt{\sum_{h=1}^L N_h^2 \cdot (1 - f_h) \cdot \frac{S_h^2}{n_h}}(\hat{X}) < X < \hat{X} + z \sqrt{\sum_{h=1}^L N_h^2 \cdot (1 - f_h) \cdot \frac{S_h^2}{n_h}}(\hat{X})\right) = \beta$$

En segundo lugar, considerando que a mostra estratificada se obtén seleccionando una mostra aleatoria simple con reempazamento en cada estrato de forma independente, a mostraxe denomínase mostraxe aleatoria estratificada con reempazamento:

Donde $P_{hi}=1/N_h$ é a probabilidade unitaria de selección da unidade u_{hi} para a mostra \tilde{X}_h de n_h unidades, obtidas de entre as N_h unidades do estrato h-ésimo.

A expresión do estimador é a mesma con reempazamento que sen reempazamento:

$$\hat{\theta}_{HH} = \sum_{h=1}^L \sum_{i=1}^{n_h} \frac{Y_{hi}}{n_h P_{hi}} = \sum_{h=1}^L \sum_{i=1}^{n_h} \frac{Y_{hi}}{n_h/N_h} = \hat{\theta}_{HT}$$

Como en tódolos estratos se realiza unha mostraxe aleatoria simple con reempazamento, o estimador de Hansen e Hurwitz (empregado con reempazamento) en cada estrato coincide co estimador de Horwitz e Thompson (empregado sen reempazamento), polo que os estimadores na mostraxe aleatoria con reempazamento coincidirán cos da mostraxe sen reempazamento.

En canto a varianza do estimador, será a seguinte:

$$V(\hat{X}_{st}) = V\left(\sum_{h=1}^L \hat{X}_h\right) = \sum_{h=1}^L V(\hat{X}_h) = \sum_{h=1}^L N_h^2 \cdot \frac{\sigma_h^2}{n_h}$$

- Mostraxe aleatoria por conglomerados: Divídese a poboación en varios grupos de características parecidas entre eles e logo analízanse completamente algúns dos grupos, descartando os demais. Dentro de cada conglomerado existe unha variación importante, pero os distintos conglomerados son parecidos. Requírese dunha mostra máis grande, pero soe simplificar a recollida de mostrase. Frecuentemente os conglomerados aplícanse a zonas xeográficas.
- Mostraxe mixta: Cando a poboación é complexa, calquera dos métodos descritos pode ser difícil de aplicar, nestes casos aplícase unha mostraxe mixta que combina dous ou máis dos anteriores métodos sobre distintas unidades da enquisa.

2. Estatística laboral: Enquisa de Poboación Activa

2.1. Introducción a Enquisa de Poboación Activa

A importancia das Estatísticas do Traballo deriva de que se encargan de describir o tamaño, estrutura, características, resultados e contribucións dos participantes no mercado de traballo e como cambian ao longo do tempo.

Segundo a International Labour Organization (2013), dende o punto de vista económico, estas estatísticas son útiles para analizar, avaliar e monitorizar o xeito no que a economía está funcionando e a eficacia das políticas económicas actuais e de longo prazo. Desde un punto de vista social, son útiles para alcanzar o traballo digno, a través de políticas e programas para a creación de emprego, formación e reconversión profesional, e asistencia aos grupos vulnerables, incluíndo os mozos, os anciáns, mulleres, etc, na procura e garantía de emprego digno .

A Organización Internacional do Traballo (OIT) propón normas internacionais sobre estatísticas do traballo que son de dous tipos: convenios e recomendacións. Foron adoptadas dúas convencións, a primeira referida ás estatísticas de salarios e horas de traballo, adoptada pola OIT en 1938, a cal ratificaron 34 países; e a segunda, que substitúe á anterior, foi aprobada en 1985. Establece directrices mínimas que abranguen tódolos ámbitos de estatísticas laborais básicos. Esta convención foi

ratificada por 48 países. En 1985 a Conferencia Internacional do Traballo tamén aprobou unha recomendación que ofrece orientación adicional en relación á frecuencia de recollida de datos e ás desagregacións recomendadas das estatísticas.

A poboación economicamente activa comprende todas as persoas de ambos sexos que proporcionan a oferta de traballo para a produción de bens e servizos durante un período de tempo de referencia indicado (OIT, 2013). O principal obxectivo da recollida de datos na poboación economicamente activa é proporcionar información básica do tamaño e da estrutura da forza de traballo. Os datos recollidos en diferentes momentos de tempo proporcionan unha base para seguir as tendencias actuais e os cambios no mercado de traballo e da situación do emprego. Estes datos, complementados coa información sobre outros aspectos da Economía, proporcionan unha base para a avaliación e análise das políticas macroeconómicas dun lugar.

Polo tanto, as estatísticas sobre a poboación economicamente activa son esenciais para o deseño e avaliación de políticas públicas xerais que pretenden promover e crear emprego. Ademais tamén poden servir para unha variedade de propósitos analíticos: os datos poden ser usados para explicar o crecemento pasado dunha economía, estudar os factores demográficos e socioeconómicos que afectan o tamaño e a composición da forza de traballo, ou para facer proxeccións da poboación economicamente activa e dos seus compoñentes como base para a planificación socioeconómica.

En canto ó deseño mostral, segundo Hussmanns, R., Mehran, F., Verma, V. (1990), para as enquisas a fogares, a mostra adoita conter un número de fases. O primeiro paso implica a selección de unidades primarias de mostraxe que son, normalmente, as unidades de tamaño relativamente grande (áreas ou municipios), entre elas deberán abordar a poboación obxecto de estudo. O marco de unidades primarias de mostraxe que comprende toda a poboación da investigación é chamada unidade primaria de mostraxe. Unha vez que unha mostra de unidades primarias de mostraxe foi seleccionada a partir da estrutura primaria de mostraxe, aplícase o mesmo á toma de mostras dentro das unidades seleccionadas, na segunda fase, e así nas sucesivas etapas se as houbera.

2.2 A Enquisa de Poboación Activa en diferentes países

Co fin de analizar os diferentes enfoques aplicados na realización da EPA a nivel mundial, farase unha comparación dos métodos que se empregan en diferentes países do mundo para obter indicadores que proporcionen información sobre o mercado de traballo.

Temos que ser conscientes de que existen países moi difíciles de comparar, xa sexa entre outros conceptos, por exemplo, pola diferenza no volume de poboación (polo que haxa que empregar mostras con máis individuos en lugares con máis poboación, para conseguir un reflexo mellor das características da mesma).

Debido as diferentes circunstancias, cada país terá unhas prioridades a ter en conta, polo que o seu método de obtención de indicadores de medición do mercado laboral e as características de ditos indicadores, dependerán das prioridades do país que estarán moi vinculadas as características do mesmo.

• Unión Europea (UE)

O organismo común para tódolos membros é o EUROSTAT (Oficina de Estatística das Comunidades Europeas), que no ámbito estudado, da Enquisa de poboación activa, encárgase de homoxeneizar e coordinar as estatísticas dos diferentes países integrados.

Os conceptos e definicións empregados na enquisa baseanse nas “Recomendacións da 13ª Conferencia Internacional de Estatísticas do Traballo”, convocada en 1982 pola Organización Internacional do Traballo.

Segundo EUROSTAT (2009) e EUROSTAT (2013), para mellorar aínda máis a comparabilidade dentro da UE, o Regulamento do 1897/2000, da unha definición máis precisa de desemprego e de emprego.

Definen as persoas empregadas como:

- As que teñen de 15 anos e máis (16 anos e máis en España, Italia, Reino Unido e Suecia).
- Que durante a semana de referencia realizaran o traballo, aínda que sexa por unha hora á semana, para o aumento salarial, beneficios ou ganancia familiar.

- Que non estaban traballando pero tiñan un emprego ou negocio do que estaban temporalmente ausentes debido a, por exemplo, enfermidade, vacacións, conflito laboral ou educación e formación.

Definen os desempregados como persoas:

- En idade de 15 a 74 (en España, Italia, Suecia, Reino Unido, Eslovenia e Noruega:16-74)
- Que estaban sen traballo durante a semana de referencia, pero dispoñibles para traballar.
- Que busquen activamente traballo nas últimas catro semanas ou que xa houberan atopado un traballo para comezar dentro dos próximos tres meses.

As persoas inactivas defínenas como aqueles que non se clasifican nin como empregados nin como desempregados.

Polo tanto, nos países membros da Unión Europea a metodoloxía de analizar o mercado laboral é similar, seguindo as orientacións do EUROSTAT.

No anexo 1 (Información que se ten en conta a hora de facer a EPA nos distintos países da UE, no ano 2011) apreciamos en base a algúns aspectos (participación, tamaño da mostra, índice de mostraxe, ratio de resposta, unidades primarias e secundarias e tipo de mostraxe) como realizan a EPA nos distintos países da UE.

Vemos como todos os países realizan unha mostraxe de tipo aleatoria, loxicamente, ó ser feita, a Enquisa de Poboación Activa, polos respectivos organismos estatais de cada país e dada a importancia que ten para a sociedade os resultados obtidos do estudo, os métodos non probabilísticos xa non se barallan debido a pouca representatividade que implican. Dentro dos probabilísticos, o máis empregado pola maioría dos países é o estratificado en dúas etapas (usado tamén en España). Este último método resulta ideal para o ámbito que queremos estudar, debido a que no país temos unha poboación heteroxénea en canto á característica que estudamos, que é o mercado de traballo, e esta mostraxe divídenos a poboación en subpoboacións o máis homoxéneas posibles non solapadas (é o que chamamos estratos). A mostra estratificada obtense seleccionando elementos de cada un dos estratos; a aplicación desta mostraxe vai axudar a que a mostra sexa representativa e tamén a diminuír o

erro de mostraxe, e así a mostra total representará a tódolos grupos da maneira máis eficaz posible.

A participación inflúe en gran medida no ratio de resposta, xa que en xeral, os países onde a participación na enquisa é obrigatoria, evidentemente, os ratios de resposta son maiores (en todos os países con participación obrigatoria, o ratio de resposta é superior o 80%, excepto Malta con un 69.2 % e Bélxica con un 67.4%).

Os índices de mostraxe máis elevados téñenos Irlanda (3%), Luxemburgo (3.6%) e Malta (2.25%), estes países ao ter menos poboación necesitan unha mostra proporcionalmente máis grande para conseguir o mesmo nivel de representatividade que outros países con máis poboación.

• Estados Unidos

Segundo a información recollida na United States Census Bureau (2013), este país para medir o seu mercado de traballo válese de varias enquisas e programas:

- *Current Population Survey (CPS)*: Enquisa de Poboación Actual.

É unha enquisa por mostraxe, de alta calidade, da poboación de 16 a máis anos; levada a cabo cada mes pola Oficina do Censo para a Oficina de Estadísticas Laborais. Proporciona datos precisos e oportunos que son necesarios para a toma de decisións económicas rápidas pola Administración, o Congreso e a Xunta da Reserva Federal.

A enquisa amplíase e mellórase ó longo dos anos, pero os conceptos básicos de emprego e desemprego que foron revisados periódicamente por comisións de alto nivel, mantivéronse practicamente inalterados.

A CPS emprega unha mostra de aproximadamente 60000 fogares ocupados. O traballo lévase a cabo durante a semana que inclúe o día 19 do mes e as preguntas refírense a semana anterior (é dicir, a que inclúe o día 12 do mes). Os fogares dos cincuenta estados e o Distrito de Columbia atópanse na enquisa durante catro meses consecutivos, saen fóra oito, e logo volven outros catro meses antes de saír da mostra de forma permanente.

Está especificamente deseñado para ser a fonte oficial das estimacións mensuais de emprego e desemprego nos Estados Unidos.

Segundo nos di Sue Zimmerman, Tamara (2012), na CPS emprégase o método de reiteracións sucesivas para a estimación da varianza, o cal analizaremos máis adiante. Fay e Train en 1995 estenderon a teoría básica das replicacións para que fose aplicable á CPS.

- *American Community Survey (ACS)*: Enquisa sobre a Comunidade Estadounidense.

É a maior enquisa de fogares nos Estados Unidos. A ACS estuda unha mostra nacional de unidades de vivenda tanto ocupadas como desocupadas, e o grupo de institucionalizados. A mostra deseñouse para proporcionar de un só ano as estimacións das características da vivenda e as características socioeconómicas para a nación, todos os estados, e as zonas con poboación de 65000 ou máis habitantes. Proporciona estimacións de 3 anos, para áreas xeográficas con una poboación de 20.000 ou máis habitantes. Para áreas con unha poboación inferior a 20.000, estarán dispoñibles as estimacións de cinco anos.

En cambio, a CPS examina unha mostra nacional dos fogares e do grupo de non institucionalizados de aproximadamente 824 áreas de mostra. A mostra deseñouse principalmente para producir estimacións das características da man de obra da poboación civil non institucionalizada de 16 anos de idade ou máis, para a nación e todos os estados.

A gran diferenza é que a ACS inclúe na súa enquisa poboación institucionalizada (por exemplo: prisións, centros de tratamento residencial e centros de enfermería) e non institucionalizada (por exemplo, colexios ou residencias universitarias e cuarteis militares), mentres que a CPS só inclúe os non institucionalizados.

Outra diferenza entre estas enquisas é que a ACS entrevista a cada persoa da unidade de vivenda no día da entrevista se está vivindo ou leva vivindo alí máis de dous meses, independentemente de se teñen ou non que manter unha residencia habitual noutro lugar. A CPS entrevista a todas as persoas que se aloxen na vivenda no momento da entrevista e se a consideran vivenda habitual ou que non teñen vivenda habitual. Tamén inclúe as persoas temporalmente ausentes que consideran a vivenda como residencia habitual.

Polo tanto, O CPS e a ACS, ambos proporcionan características socioeconómicas que o outro lle falta por completo ou non teñen o mesmo nivel de detalle; o ACS, con un tamaño de mostra máis grande, pode soportar tabulacións máis detalladas que o CPS. E o CPS ten datos anuais dispoñibles só para os estados e as grandes áreas metropolitanas.

- *Local Area Unemployment Statistics (LAUS) Program:* Programa de Estatísticas de Desemprego da Área Local.

Produce estimacións mensuais e anuais de emprego, desemprego e a tasa de desemprego de máis de 7.000 áreas xeográficas. Os datos da CPS son unha base clave para a metodoloxía do programa LAUS. A CPS e as estimacións do programa LAUS proporcionan unha serie consistente para o emprego e os datos do desemprego a nivel nacional e estatal.

- *Current Employment Statistics (CES) Program:* Programa de Estatísticas de Emprego actuais.

Fonte de información sobre emprego, horas traballadas e os ingresos por hora e semanais segundo o informado por unha mostra de cerca de 400.000 establecementos. Emprego, horas traballadas e os datos de ingresos baséanse en informes de nómina. Este estudo estima o número e as características dos traballos retidos, non o número de persoas empregadas.

- *Survey of Income and Program Participation (SIPP):* Enquisa de Ingresos e Participación en Programas.

É unha enquisa lonxitudinal. É útil principalmente para examinar a dinámica do cambio no emprego (incluíndo períodos de desemprego) para os mesmos individuos no tempo.

- *Unemployment Insurance (UI) Administrative Records:* Seguro de Desemprego. Rexistros Administrativos.

As estatísticas sobre as persoas que reciben beneficios de seguro de desemprego recóllense como un subproduto dos programas de seguro de

desemprego. Os traballadores que perden os seus empregos e están cubertos por estes programas soen presentar demandas que serven como aviso de que están comenzando un período de desemprego.

Estos informes elabóranse a partir de tres fontes principais:

- As enquisas de fogares, como as CPS, A ACS e a SIPP; proporcionan datos que pertencen os individuos, as familias e os fogares, e refírense ó lugar donde residen (en oposición o lugar de traballo).
- As enquisas de establecementos, como o CES, proporcionan datos dos empregadores ou establecementos que pertencen a postos de traballo e refírense os postos de traballo donde se encontren.
- Os datos basados na información do Seguro de Desemprego de rexistros administrativos, refírense a persoas que son beneficiarias de desemprego según os requisitos do programa, incluíndo os que presentan reclamos e relaciónanse co lugar de cobertura do seguro do establecemento.

As estimacións de emprego e desemprego dun programa en particular poden diferir das de outros programas debido ás variacións nas definicións, de cobertura, métodos de recolección, períodos de referencia, etc. As distintas estimacións de emprego e desemprego, en xeral, complementáanse entre sí.

(Anexo 2: Situacións nas que resulta apropiada cada enquisa nos Estados Unidos)

• Brasil

A través do Insitituto Brasileiro de Xeografía e Estatística publícanse entre outras, investigacións relativas ó emprego. Máis concretamente, a fonte principal que nos aporta información do mercado de traballo en Brasil é a “*pesquisa mensal de emprego*”, unha investigación mensual de emprego, que realiza estimacións mensuais sobre a forza de traballo.

A PME (*pesquisa mensal de emprego*) é unha enquisa a fogares que investiga características da poboación que reside en áreas de cobertura urbanas-

metropolitanas. O tema fundamental da PME é o traballo, consistindo na investigación das características sociodemográficas e educacionais co obxectivo de conseguir unha mellora na comprensión da forza de traballo (IBGE, 2007).

De igual modo que ocorreu en España, a investigación mensual de emprego en 2001, modificouse co fin de adaptarse as recomendacións da Organización Internacional do Traballo (OIT); sendo os principais cambios metodolóxicos introducidos co fin de ampliar as investigacións para comprender mellor a poboación activa e a poboación en busca de traballo. Tamén se produciron cambios nos procedementos para a recolección de datos da mostra coa recollida electrónica de datos e tamén cambios no relativo a expansión da mostra.

A cobertura xeográfica da investigación mensual de emprego céntrase en áreas metropolitanas de Recife, Salvador, Belo Horizonte, Río de Xaneiro, Sao Paulo e Porto Alegre.

A investigación mensual de emprego, lévase a cabo a través dunha mostra aleatoria de fogares, deseñada para garantir a representatividade dos resultados para os niveis xeográficos onde se produce a investigación. As persoas incluídas na enquisa son todas as residentes, excepto as persoas que viven en embaixadas, consulados e legacións e as persoas institucionalizadas que viven en fogares colectivos de institucións como o cuartel militar, en instalacións militares, reclusos nas cárceres, reformatorios, internos en escolas, hospitais, asilos, orfanatos, relixiosos nos conventos, etc.

As definicións da investigación mensual de emprego en Brasil coinciden en gran medida cas que nos proporciona o INE en España.

A poboación economicamente activa corresponde a parte da poboación de dez anos ou máis que tiñan ou estaban buscando algún traballo na semana de referencia (IBGE, 2001)

Aínda que resulta sorprendente unha idade tan nova (máis de dez anos), a partir da cal se considera a poboación para catalogala nun posto do mercado de traballo ou noutro, cabe mencionar, que en Brasil, na maioría dos grupos xóvenes (de 10 a 14 anos e de 15 a 17), estase producindo unha diminución das taxas de actividade a medida que nos acercamos á actualidade grazas a loita contra o traballo infantil e as políticas impostas neste ámbito.

As persoas que declararan ter un traballo remunerado e que non o exerceran, durante polo menos unha hora completa na semana de referencia, invéstigase o

motivo de non exercer o traballo e o tempo no que estiveron afastados do traballo que tiñan. O motivo do afastamento resulta determinante para a consideración da persoa como ocupada ou non, na semana de referencia.

Ás persoas ocupadas durante a semana de referencia, identifícanlle cantos traballos tiñan na semana de referencia e cal foi o traballo principal, caracterizando este último.

As persoas ocupadas clasifícanse en empregados, por conta propia, empregadores e traballadores non remunerados do fogar que era o seu propio empregador.

A investigación mensual de emprego brasileira outórgalle gran importancia ó “ingreso mensual do traballo”, para o cal fai variedade de estatísticas que se poden visualizar no Instituto Brasileiro de Xeografía e Estatística. Segundo este, considérase como un ingreso mensual de traballo:

- Para os empregados: Unha remuneración mensual bruta que normalmente tería dereito ou cando o rendemento é variable, o salario promedio mensual para o mes analizado.
- Para os empresarios e traballadores autónomos: O retiro mensual ou cando o rendemento é variable, o retiro mensual promedio para o mes analizado.

En canto a realización da mostra, tamén é moi similar a que se emprega no noso país. A PME utiliza unha mostra probabilística de fogares, estratificada e por conglomerados mediante mostraxe en dúas etapas, para cada área metropolitana cuberta na enquisa.

Nos municipios e pseudomunicipios (conxunto de municipios pequenos segundo o Censo Demográfico), cada un dos estratos corresponde a unha selección independente, asegurando así a propagación da mostra pola área metropolitana.

Dentro de cada municipio ou pseudomunicipio faise unha selección das unidades primarias de mostraxe (as seccións censais) e das unidades secundarias posteriormente (unidades de fogares).

A selección das unidades primarias de mostraxe lévase a cabo mediante unha mostraxe sistemática con probabilidade proporcional ó número total de ocupados de fogares particulares, obtidos polo Censo.

A selección das unidades secundarias de mostraxe (selección dos fogares) faise mediante mostraxe sistemática simple. A PME conta con unha serie de aspectos dunha mostraxe autoponderada dentro de cada área metropolitana (dependendo

do crecemento natural da industria, o número de fogares seleccionados, vai variar).

A fracción da mostra, obtense dividindo o número total de unidades de aloxamento polo número desexado dos fogares na mostra, fixando en 16 o número de fogares seleccionados en cada sector, chegouse o número de sectores na mostra.

A enquisa é distribuída polas catro semanas de referencia do mes, polo que o resultado mensual é obtido polo conxunto da información das catro semanas de referencia.

A recolección dos datos segue unha metodoloxía, na cal cada fogar seleccionado permanece catro meses consecutivos sendo investigado, oito meses sen ser investigado, e despois deste período, é investigado novamente por catro meses máis e finalmente excluída da mostra. Se durante o período que a unidade familiar permanece na mostra, a familia cambia de enderezo e outra familia pasa a ocupar o fogar, a información será obtida cunha nova familia no período restante.

A mostra mensual da PME é subdividida en oito grupos de rotación. Cada mes o 25% da mostra de fogares substitúese, seguindo un esquema de grupos de rotación e paneis.

Este método de rotación persigue obter vantaxes nas variacións de estimacións mes a mes, xa que tres cuartas partes da mostra mantéñense en meses consecutivos e variacións de estimacións de diferenza ano a ano xa que a metade da mostra repítese no mesmo mes en anos consecutivos. De este modo, tense para calquera mes, fogares entrevistados dende a primeira ata a oitava vez.

(Anexo 3: Porcentaxe de fogares en común, na mostra, en Brasil, entre pares de meses)

• Canadá

Na economía canadense, o indicador que proporciona estimacións do mercado laboral é a “Labour Force Survey”, a enquisa de poboación activa, que posúe multitude de similitudes coa EPA en España.

Xunto coa EPA, para conseguir unha maior e máis completa información do mercado de traballo, encóntranse a Enquisa de Emprego, Nóminas e Horas e as Estatísticas de Seguro de Emprego.

Segundo a información recollida en Statistics Canada (2013), a EPA cubre á poboación civil, non institucionalizada de 15 anos ou máis de idade. Lévese a cabo en todo o país, excluíndose da cobertura da enquisa as persoas que viven en reservas e outros asentamentos aboríxes nas provincias, os membros das Forzas Armadas a tempo completo e os grupos institucionalizados de poboación. (Estes últimos grupos representan unha exclusión de menos do 2% da poboación canadense de 15 anos ou máis).

O cuestionario actual da EPA introduciuse en 1997 co fin de abordar ausencias nos datos existentes, mellorar a súa calidade e facer un uso maior do “Computer Assisted Interviewing” (CAI).

En canto á mostraxe, seguindo a información proporcionada na web de Statistics Canada (2013), trátase dunha enquisa por mostraxe cun deseño transversal. A EPA emprega unha mostra de probabilidade baseada nun estratificado en multietapa. Cada provincia está dividida dentro de un gran estrato xeográfico. A primeira etapa de mostraxe consiste en seleccionar áreas xeográficas máis pequenas, chamadas agrupacións, dentro de cada estrato. A segunda etapa consiste na selección das vivendas dentro de cada grupo seleccionado.

A EPA utiliza un deseño da mostra de panel rotativa coa finalidade de que as vivendas seleccionadas permanezan na mostra da EPA durante seis meses consecutivos. Cada mes ao redor dun sexto das vivendas incluídas na mostra atópanse no seu primeiro mes da enquisa, un sexto están no seu segundo mes da enquisa, e así sucesivamente. Cada un dos seis grupos de rotación pode ser utilizado como unha mostra representativa de por sí.

Segundo “Statistics Canada (2013)”, dende xullo de 1995, o tamaño da mostra da EPA mensual foi de aproximadamente 54000 fogares (100000 persoas, aproximadamente). A resposta a enquisa é obrigatoria, sendo a semana de referencia, normalmente, a que contén o día 15 do mes.

2.3. A Enquisa de Poboación activa en España

No noso país, en liñas xerais, todos os autores coinciden na definición da EPA que nos ofrece o Instituto Nacional de Estatística: “A EPA é unha investigación por mostraxe de periodicidade trimestral realizada polo INE, dirixida á poboación que reside en vivendas familiares do territorio nacional e cuxa finalidade é averiguar as características de dita poboación en relación co mercado de traballo.” (INE, 2008)

Polo tanto, a EPA pretende coñecer a actividade económica no relativo á súa compoñente humana. Persigue ofrecer datos das principais categorías poboacionais en relación co mercado de traballo (ocupados, parados, activos e inactivos) detallados para o conxunto nacional. Para as comunidades autónomas e as provincias ofrécese información sobre as principais características co grado de desagregación que permite o coeficiente de variación dos estimadores. Posibilita confeccionar series temporais homoxéneas de resultados, comparables no tempo e homologables cos de enquisas similares de outros países.

As fontes estatísticas das que dispoñiamos (o Censo, a Enquisa de Salarios, o Paro rexistrado, Afiliación a Seguridade Social,...) non son óptimas para describir especificamente o mercado laboral.

O censo ten unha periodicidade moi longa, co cal impídenos coñecer a situación en períodos intercensais; ademais de que os datos son obtidos por autoinscripción (o entrevistado cumprimenta él mesmo o cuestionario), existindo en moitos casos dificultades de interpretación por parte do informante.

A enquisa de salarios só nos aporta información da poboación asalariada.

Debido á variabilidade da normativa legal que rexe o Paro Rexistrado e a Afiliación á Seguridade Social, presentan dificultades na obtención de series homoxéneas e só nos proporcionan información dunha parte da poboación estudada (parados rexistrados ou ocupados afiliados á Seguridade Social).

Por causa desta variedade de fontes, existe dende hai anos un debate: por un lado, os que opinan que ante a existencia dos datos de afiliación á Seguridade Social e das cifras do paro rexistrado non é necesaria a EPA; e por outro lado, temos os que

defenden a utilidade que nos proporciona a EPA. Entre estes últimos e para xustificar a súa defensa, o Presidente do INE (Gregorio Izquierdo) nun artigo do día 8/03/2013 no xornal “Cinco Días” ofrécenos argumentos que xustifican a veracidade da EPA: “Os datos de afiliación non serían comparables co resto dos países. Trátase de rexistros complementarios e non substitutivos. A utilidade da EPA xustifícase porque non se limita a cuantificar á poboación ocupada, senón que aborda as súas características, como o nivel educativo, por non dicir a información que facilita sobre os inactivos.”

“Cando existe unha estatística tan fiable, veraz e contrastada, o dato que da a EPA é o que realmente hai (...) Existen incentivos a rexistrarse como parado como é a posibilidade de encontrar un emprego, cobrar a prestación ou recibir formación gratuíta. E en tempos de crises como os actuais o que pode ocorrer é que haxa persoas que aínda estando desempregadas non lles compense rexistrarse. A EPA a este respecto é máis exhaustiva. Non só cuestiona ó enquisado se está rexistrado como parado, senón que lle pregunta se realiza unha búsqueda activa de emprego. Non existen incentivos a mentir ó entrevistador.”

Co obxectivo de expoñer a teoría de mostras na EPA en España, e para comprendela mellor situándonos no ámbito que se emprega na enquisa, mencionaranse as definicións utilizadas e as taxas que se derivan de ditas definicións.

- **ÁMBITO DA ENQUISA:**

Desglosarémo-lo en ámbito poboacional, ámbito xeográfico e ámbito temporal para comprender o entorno que abarca a enquisa, segundo o INE (2012):

- Ámbito poboacional: A Enquisa vai dirixida o conxunto da poboación que reside en vivendas familiares (as utilizadas todo o ano ou a maior parte del como vivienda habitual). Exclúense da investigación os fogares colectivos (hospitais, hoteis, conventos, etc). Inclúense as familias que formando un grupo independente residan en estos establecementos (directores dos centros, conserxes, porteiros, etc).
- Ámbito xeográfico: A enquisa realízase en todo o territorio nacional.
- Ámbito temporal: A EPA é unha enquisa continua con periodicidade trimestral, realizándose as entrevistas o longo das trece semanas do trimestre. O período

de referencia dos resultados da enquisa é o trimestre. O período de referencia da información recollida, como normal xeral, é a semana anterior (de luns a domingo) da data na que se realiza a entrevista.

- **DEFINICIÓNS:**

As definicións empregadas na EPA están baseadas nas recomendacións da Organización Internacional do Traballo (OIT) aprobadas na XIII e XVI Conferencia Internacional de Estadísticas do Traballo (Ginebra, 1982 e 1988, respectivamente).

Partindo da información que nos proporcionan Alba Fernández, V. e Muñoz Vázquez, A. (2000) e o INE (2008), deriváanse as seguintes definicións:

- Poboación economicamente activa: O conxunto de persoas de 16 anos ou máis que no período de referencia dado, suministran man de obra para a produción de bens e servizos económicos ou que están dispoñibles e realizan xestións para incorporarse a dita produción.
- Poboación ocupada: Aquelas persoas de 16 anos ou máis que durante a semana de referencia tiveron traballo por conta allea (asalariados) ou exerceron unha actividade por conta propia, podendo estar durante esa semana:
 - Traballando polo menos unha hora a cambio dun soldo, salario, beneficio empresarial ou garantía familiar, en metálico ou en especie.
 - Con emprego pero sin traballar: ausentes do seu traballo pero esperando poder incorporarse o término da contixencia que orixinou a ausencia.

Hai que ter en conta que:

1. Non son ocupados as persoas que coidan do seu fogar sen remuneración, as que prestan servizos sociais ou de carácter benéfico non remunerado e, en xeral, todas as que exerzan actividades sin un fin lucrativo.
2. Son ocupados os estudantes que traballan a cambio dunha remuneración e os aprendices que reciben unha remuneración en metálico ou en especie.

Dentro da ocupación, a “*situación de subemprego*” existe cando a ocupación que ten unha persoa é inadecuada respecto a determinadas normas ou a outra ocupación posible, tendo en conta a súa calificación profesional. Na EPA considérase que “son subempregados todos os ocupados que ou ben traballan a tempo parcial por non poder encontrar un traballo a xornada completa e están buscando outro emprego, ou ben están afectados por un expediente de regulación de emprego, con suspensión ou redución de xornada.”

- Poboación parada: Considéranse paradas todas as persoas de 16 anos ou máis que durante a semana de referencia estiveran:
 - Sen traballo: non tiveron un emprego por conta allea ou por conta propia durante a semana de referencia.
 - En busca de traballo: tomaron medidas concretas para buscar un traballo por conta allea ou fixeron xestións para establecerse pola súa conta durante o mes precedente.
 - Dispoñibles para traballar: en condicións de comezar a facelo no prazo de dúas semanas a partir da data da entrevista.

Os parados subdivídense en parados que buscan o seu primeiro emprego e parados que xa traballaron anteriormente.

- Poboación economicamente inactiva: Comprende a todas aquelas persoas de 16 anos ou máis, non clasificados como ocupados ou parados durante a semana de referencia. Comprende as seguintes situacións:
 - Persoas que se ocupan do seu fogar: sen exercer ningunha actividade económica adócanse a coidar sin remuneración dos seus propios fogares; por exemplo, as amas de casa e outros familiares que se ocupen do coidado da casa e dos nenos.
 - Estudantes: persoas que, sen exercer ningunha actividade económica, reciben unha instrución sistemática en calquer grado de educación.
 - Xubilados ou prexubilados: persoas que tiveron unha actividade económica anterior e que pola súa idade e outras causas, abandonárona.

Sendo o seu medio de vida as pensións obtidas pola súa actividade anterior.

- Persoas que reciben unha pensión distinta da de xubilación.
- Persoas que realizan sen remuneración traballos sociais, actividades de tipo benéfico, etc.
- Incapacitados para traballar.
- Outra situación: persoas que sen exercer ningunha actividade económica, reciben axuda pública ou privada e todas aquelas que non estean incluídas en ningunha das categorías anteriores (por exemplo, os rentistas).
- Poboación contada a parte: A partir do 2002 coa desaparición do servizo militar obligatorio, deixou de existir esta categoría poboacional. Anteriormente a esa data, os varóns que cumprían o servizo militar, non se lles incluía entre os activos nin entre os inactivos, independentemente de que na semana de referencia traballasen ou non.

(Anexo 4: Esquema de clasificación da poboación de 16 anos ou máis)

• **TAXAS:**

Partindo das definicións anteriores, podemos derivar as seguintes taxas, das cales, os seus resultados clarificarán a situación do mercado laboral:

Taxa global de actividade	$\frac{\text{Nº total de activos}}{\text{Poboación total}}$
Taxa específica de actividade	$\frac{\text{Nº de activos dun intervalo de idade concreto}}{\text{Poboación total de ese intervalo de idade}}$
Taxa de actividade	$\frac{\text{Total de activos}}{\text{Poboación de 16 anos ou máis}}$
Taxa global de emprego	$\frac{\text{Nº total de ocupados}}{\text{Poboación total}}$

Taxa específica de emprego	$\frac{\text{Nº ocupados de un intervalo de idade determinado}}{\text{Poboación total de ese intervalo determinado}}$
Taxa de paro	$\frac{\text{Nº de parados}}{\text{Nº de activos}}$
Taxa específica de paro	$\frac{\text{Nº de parados de un intervalo de idade determinado}}{\text{Nº de activos de ese intervalo de idade}}$
Taxa de asalarización	$\frac{\text{Nº de asalariados}}{\text{Nº total de ocupados}}$
Taxa de temporalidade	$\frac{\text{Nº de asalariados con contrato temporal}}{\text{Nº total de asalariados}}$
Taxa de traballo a tempo parcial	$\frac{\text{Nº de ocupados a tempo parcial}}{\text{Nº total de ocupados}}$

Fonte: elaboración propia a partir das definicións empregadas na EPA, elaboradas coa información proporcionada de Alba Fernández, V. e Muñoz Vázquez, A. (2000) e o INE (2008).

Anteriormente foi exposta polo miúdo a Teoría de Mostras, agora, partindo de dita teoría, explícase o seu procedemento de aplicación na EPA, en España, a cal se fai do seguinte modo, como nos indica V. e Muñoz Vázquez, A. (2000) e o INE (abril, 2012):

A EPA encóntrase enmarcada dentro do deseño da Enquisa Xeral da Poboación (modelo utilizado polo INE para as enquisas dirixidas os fogares).

- O *MARCO*: Está baseado na división administrativa oficial con 17 Comunidades Autónomas e 50 provincias, que a súa vez se subdividen en municipios. O INE xunto cos concellos, realiza unha posterior subdivisión dos distritos en seccións censais.

- *TIPO DE MOSTRAXE*: Mostraxe bietápica con estratificación das unidades da primeira etapa.

As unidades da primeira etapa están constituídas polas seccións censais. A mostra de seccións permanece fixa indefinidamente cas excepcións seguintes: saen da mostra aquelas seccións nas que xa se visitaron todas as vivendas enquisables; e cando no

proceso de actualización do seccionado a algunhas seccións lles corresponda saír da mostra, ben polos cálculos probabilísticos ou por cambios na afixación por estratos.

As seccións que saen da mostra son substituídas por outras seccións seleccionadas aleatoriamente.

As unidades da segunda etapa están formadas polas vivendas familiares principais e os aloxamentos fixos.

- *ESTRATIFICACIÓN DAS UNIDADES DE MOSTRAXE*: As unidades da primeira etapa estratifícanse atendendo a un dobre criterio:

1. Criterio xeográfico (de estratificación): As seccións agrúpanse en estratos dentro de cada provincia, de acordo coa importancia demográfica do municipio o que pertencen.
 2. Criterio socioeconómico (de subestratificación): As seccións censais agrúpanse en subestratos dentro de cada un dos estratos, según as características socioeconómicas das mesmas.
- *ESTRATOS*: Para chegar á formación dos estratos considéranse os seguintes tipos de municipios: municipios autorepresentados (deben ter sempre seccións na mostra: capital da provincia, municipios que teñen un número de habitantes tal que na afixación proporcional dentro da provincia correspóndenlle ó menos 12 seccións na mostra, municipios que tendo unha situación demográfica destacada dentro da provincia non hai outros similares cos que agrupalos) e municipios correpresentados (forman parte dun grupo de municipios demográficamente similares e que son representados en común).

En rasgos xerais, os estratos teóricos considerados serán:

Estrato1: Municipio capital de provincia.

Estrato 2: Municipios autorrepresentados.

Estrato 3: Outros municipios autorrepresentados, importantes en relación ca capital ou municipios maiores de 100000 habitantes.

Estrato 4: Municipios entre 50000 e 100000 habitantes.

Estrato 5: Municipios entre 20000 e 50000 habitantes.

Estrato 6: Municipios entre 10000 e 20000 habitantes.

Estrato 7: Municipios entre 5000 e 10000 habitantes.

Estrato 8: Municipios entre 2000 e 5000 habitantes.

Estrato 9: Municipios menores de 2000 habitantes.

- *SUBESTRATOS*: No proceso de formación dos subestratos, dentro de cada estrato consideráronse dous grupos de seccións:
 - Seccións dos estratos 7, 8 e 9: grupo de seccións pertencentes a municipios pequenos (teñen unha variabilidade relativamente pequena respecto das variables obxectivo e que está ben explicada polo territorio ó que pertencen). Asígnaselle como subestrato a comarca do municipio ó que pertencen.
 - Resto de seccións: Agrúpanse dentro dos seus estratos mediante a aplicación de técnicas de análise de conglomerados. Como son municipios moito máis grandes considerouse prioritario empregar a información auxiliar dispoñible (procedente do Censo de 2001 e da Axencia Estatal de Administración Tributaria) para formar grupos homoxéneos de seccións e mellorar así a precisión das estimacións.

As variables auxiliares utilizadas a nivel de sección foron: Porcentaxe de parados, de inactivos, de ocupados, de estranxeiros, de persoas entre 0 e 19 anos, de persoas entre 15 e 24 anos, de persoas de 65 ou máis anos, de persoas con nivel de estudos realizados 1,2 ou 3, según a clasificación do Censo de 2001 (analfabetos, sen estudos ou nivel de estudos de primeiro grado), de persoas con nivel de estudos realizados 4,5,6 ou 7 (ESO, ECB, Bacharelato, FP); de persoas con nivel de estudos realizados 8, 9 ou 10 (diplomatura, licenciatura ou doutorado).

Ademáis tomáronse 18 variables de porcentaxes de poboación ocupada da sección segundo a súa condición económica.

As variables fiscais empregadas foron: Renda total por vivenda con perceptores, renda de capital mobiliario e inmobiliario sobre renda total e renda agraria sobre renda total.

Antes do análise de conglomerados estandarízanse as variables dentro de cada estrato con media 0 e desviación típica 1 (excepto as variables: porcentaxe de

parados, de xóvenes e as variables fiscais, que se estandarizaron con desviación típica 2). Preténdese que estas últimas variables teñan unha ponderación superior ó resto.

Elimínanse do análise de conglomerados aquelas variables que representan menos do 1% do total do estrato en cada provincia.

Utilizouse un algoritmo acumulativo que obtén conglomerados xerárquicos. Pártese de tantos conglomerados como seccións e, en cada etapa, únense os conglomerados máis parecidos entre sí. Ó final todas as seccións forman un único conglomerado. Determinábase o punto intermedio do proceso de agrupación, en función do número e tamaño de conglomerados que se considere máis axeitado.

TAMAÑO DA MOSTRA: Estableceuse mediante a aplicación dun procedemento de mínima varianza para coste fixo. Partindo dun presuposto (Q) determínase o número de seccións (n) e o número de vivendas (m) que minimizan a varianza das estimacións. Utilizouse unha función de coste de tipo lineal e a expresión do coeficiente de variación para unha proporción na mostraxe de conglomerados con submostreo.

$$Q = n \cdot Q_s + n \cdot m \cdot Q_v \quad \text{con} \quad Q_s = Q_F + d \cdot Q_D$$

Q = Presuposto total

Q_s = Coste por unidade primaria (sección)

Q_v = Coste por unidade última (vivenda)

n = Número de seccións

m = Número de vivendas por sección

Q_F = Coste fixo por sección

Q_D = Coste diario do traballo de campo

d = Número de días necesarios po traballo de campo

Sendo todas as variables coñecidas menos n e m.

O coeficiente de variación para unha proporción P:

$$CV^2(\hat{P}) = \frac{V(\hat{P})}{\hat{P}^2} = \frac{1-\hat{P}}{\hat{P}} \cdot \frac{1+\delta(m-1)}{nm} = \frac{1-\hat{P}}{\hat{P}} \cdot F(\delta, m, n)$$

$$F(\delta, m, n) = \frac{1+\delta(m-1)}{nm}$$

Sendo δ o coeficiente de correlación intraclásica (para o caso da poboación activa vale 0.05).

O mínimo da expresión $CV^2(\hat{P})$ respecto das variables m e n obténse calculando o mínimo da expresión $F(\delta, m, n)$ que é independente de P .

O valor mínimo de $F(\delta, m, n)$ respecto de m e n correspondeu a $m=20$ e $n=3000$.

Coas modificacións producidas no ano 2005, estableceuse, en rasgos xerais, un tamaño mostral de 3588 seccións e 18 vivendas por sección.

Na Comunidade Autónoma de Galicia, na resolución do 10 de novembro de 2008 do Instituto Nacional de Estatística, publicouse o Convenio de colaboración entre o Instituto Nacional de Estatística (INE) e o Instituto Galego de Estatística (IGE), para a realización da enquisa de poboación activa no ámbito territorial da Comunidade Autónoma de Galicia. (BOE, 2008)

Debido a que o IGE ten encomendado no Plan Galego de Estatística a explotación dos datos sobre a actividade, ocupación e desemprego da EPA en Galicia. Co obxectivo de difundir os datos da poboación en relación coa actividade que reside na comunidade autónoma e tamén a actividade “Estudo sobre a relación coa actividade da poboación xuvenil”, o IGE, fíxolle necesario para lograr os seus obxectivos, unha ampliación da mostra da EPA na Comunidade Autónoma de Galicia. Por isto, para evitar duplicidades na EPA en Galicia e lograr atender as necesidades de información do IGE, buscando unha xestión máis eficiente dos recursos das Administracións Públicas, ambas partes acordaron establecer o Convenio de colaboración citado anteriormente.

De este modo, o INE atende a solicitude do IGE duplicando a mostra no ámbito territorial galego, pasando a ser de 468 seccións (o que equivale a unha mostra en torno a 8000 vivendas). Ademais, asígnanse estratos separados ós municipios de Santiago de Compostela e Ferrol.

O total da mostra final para o total nacional establécese en 3822 seccións.

AFIXACIÓN: Na afixación provincial tivéronse en conta as seguintes cuestións:

- Maior fiabilidade posible nos resultados nacionais.
- Disposición en cada provincia de un tamaño mínimo de mostra que permita dar estimacións.
- En cada provincia o número de seccións debe ser múltiplo de trece (facilitándose a distribución da mostra nas trece semanas do trimestre)

Adoptouse unha afixación de compromiso entre a uniforme e a proporcional, agrupando as provincias cunha importancia demográfica similar e asignándolle de 3 a 12 entrevistadores (de 39 a 156 seccións mostrais).

Dentro de cada provincia a afixación entre estratos é proporcional ó tamaño de cada un deles. Potenciáronse os estratos donde están os municipios de maior tamaño (donde a dispersión debe ser maior e o coste por entrevista menor).

Dentro dos estratos, a afixación entre subestratos é estrictamente proporcional ó tamaño (medido en número de vivendas familiares).

SELECCIÓN DA MOSTRA: Realízase de tal modo que se teñan mostrais autoponderadas dentro de cada estrato (calquera vivenda familiar ten a mesma probabilidade de ser seleccionada).

A probabilidade de selección da vivenda i , pertencente á sección x do estrato h , donde se fixaron k_h seccións, sería:

$$P(V_{ixk}) = P(S_{xh}) * P(V_{ixh} / S_{xh}) = K_h * \frac{V_{xh}}{V_h} * \frac{18}{V_{xh}} = k_h * \frac{18}{V_h}$$

$P(S_{xh})$ = probabilidade de selección da sección x do estrato h .

$P(V_{ixh} / S_{xh})$ = probabilidade de selección da vivenda i condicionada á selección da sección x .

V_{xh} = total de vivendas da sección x

V_h = total de vivendas do estrato h .

TURNOS DE ROTACIÓN: A distribución da mostra é uniforme no tempo (en cada provincia o número de seccións por semana é constante). Cada período da enquisa é de un trimestre repetíndose sucesivamente.

As seccións censais permanecen fixas na mostra indefinidamente, as vivendas familiares son renovadas parcialmente cada trimestre da enquisa nunha sexta parte das seccións (para evitar o cansancio das familias: cada vivenda permanece na mostra seis trimestres consecutivos e logo sae da mesma para ser reemplazada por outra da mesma sección).

A mostra total está dividida en seis submostras que denominamos “turnos de rotación”.

CAMBIOS PRODUCIDOS NA EPA ATA A ACTUALIDADE:

Como xa se mencionou con anterioridade, a EPA publicouse por primeira vez en 1964. A partir desta data e co fin de mellorar a calidade dos seus resultados, producíronse diversos cambios.

Dende a súa publicación ata finais de 1968 obtivéronse resultados con referencia trimestral. De 1969 a 1974 a referencia temporal pasou a ser semestral e a partir de 1975 volveu a ser trimestral.

En 1987 houbo unha modificación no cuestionario da enquisa debido a esixencia de adaptar a EPA a Enquisa de Forza de Traballo da Comunidade Económica Europea con motivo da incorporación de España a Comunidade Europea en 1986. Aquí comezou a “serie homoxénea” da EPA que se ofrece actualmente.

En 1999, a EPA pasou a ser unha enquisa continua: as entrevistas pasan a facerse ó longo das 13 semanas do trimestre (ata o momento facíanse en 12 das 13 semanas do trimestre).

En 2002, introduciuse unha nova definición operativa de paro, considerándose métodos activos de búsqueda de emprego nas catro semanas anteriores á entrevista (ata o momento, a mera inscrición nas oficinas de emprego público, válida por 3 meses bastaba para considerar parada á persoa). Empregáronse proxeccións da poboación do Censo de 1991 revisadas e introduciuse unha mellora técnica no cálculo dos resultados da enquisa (reponderación).

Este cambio afectou en gran medida xa que unha parte das persoas que ata o momento se consideraban paradas, pasaron a ser inactivas.

En 2005, según o Boletín Informativo do INE (2005):

“A EPA modifícase polas seguintes causas: para adecuarse a nova realidade demográfica e laboral do noso país (principalmente, o aumento do número de estranxeiros residentes), pola incorporación da nova normativa europea seguindo as normas da Oficina Estatística da UE(EUROSTAT) e a introducción de melloras no método de recollida.”

Polo tanto, as novidades serían:

- Actualización da base da poboación empregada:
 - Revísanse as cifras da poboación: A partir do 2001 entraron máis de 600.000 inmigrantes cada ano que é unha cifra superior en máis do triple o previsto na proxección.

(Anexo 5: Aumento da poboación estranxeira a partir de 2001)

Ata finais de 2004 empregábanse proxeccións baseadas no Censo de 1991 revisadas, agora substitúense por proxeccións baseadas no Censo de 2001, das que se dispón información de datos observados para o período 2001-2004.

- Cambio no cálculo dos factores de elevación: debido as dificultades para entrevistar a poboación inmigrante prodúcese unha modificación no cálculo dos factores de elevación consistente en que se poda reflexar mellor o peso específico dos estranxeiros en determinados sectores. Deste modo, sempre que alcancen un mínimo de representación na mostra, incorpórase o proceso de calibrado a variable nacionalidade (é dicir, mediante este proceso matemático reaxústanse os factores de elevación obtidos a partir do deseño mostral: axústase a pirámide de poboación por sexo e idade de cada comunidade autónoma, o total de poboación de cada provincia e o total de poboación por nacionalidade de cada comunidade autónoma de forma que coincidan coas cifras reais).

- Nova normativa de EUROSTAT:
 - Introdúcense novas variables: EUROSTAT especifica novas variables que hai que recoller nas enquisas de poboación activa dos Estados Membros.

A Enquisa de Forza de Traballo Comunitaria especifica que se teñen que incluír seis novas variables (se a persoa no seu emprego principal supervisa as tarefas de outros traballadores, se ten contrato con unha axencia de emprego temporal, o papel que desempeñaron as oficinas públicas de emprego na obtención do traballo principal actual da persoa entrevistada, caracterizar as horas extraordinarias e se son pagadas ou non, existencia ou carencia de servizos de coidados a persoas dependentes e a percepción ou non do salario no caso de ausencia prolongada do traballo)

- Sistema de submostras: Temos preguntas no cuestionario ás que necesariamente hai que dar información trimestral (constitúen o núcleo) e outras preguntas que son as estruturais, ás que é necesario dar información unha vez ó ano e para as que se crea o sistema de submostras. Lógrase diminuír a duración das entrevistas xa que cada familia colaboradora só responde o cuestionario extenso unha vez de seis que participa na enquisa.
- Modernización do método de recollida da información:
 - Cambios no cuestionario: Estandarízanse as entrevistas de modo que todos os entrevistadores teñen que seguir exactamente o guión que marca o cuestionario, así o cuestionario é un instrumento de medida neutral, sen posibilidade de que o entrevistador introduza variabilidades. Preténdese captar situacións especiais en relación o emprego e as horas traballadas, as novas preguntas permiten detectar mellor os empregos de moi curta duración ou os de xornada moi reducida. O efecto desta normalización é un aumento das preguntas do cuestionario.
 - Entrevistas realizadas co método CATI: O entrevistador introduce as respostas nun ordenador no que se programou o cuestionario, permitindo seguir as entrevistas cando se están facendo. A primeira entrevista ó fogar continuase realizando mediante visita persoal.

3. Medición dos erros de mostraxe: aplicación práctica

Como parte final desta análise vaise replicar o cálculo da varianza mostral para un parámetro, neste caso a proporción de parados (taxa de paro). Aplicarase este cálculo a datos procedentes da enquisa de poboación activa do terceiro trimestre de 2009, recollidos do IGE (2013). Hai que ter en conta, como xa se mencionou con anterioridade, que a mostra empregada en Galicia está duplicada respecto a outros anos polo acordo establecido entre o IGE e o INE.

Usar estes datos implica unha situación moito máis simple ca real, polo que será necesario adaptar a metodoloxía, o que nos permitirá unha comprensión maior de diferentes conceptos mencionados no traballo.

En calquera aplicación de mostraxe interesa coñecer se a nosa estimación é de calidade, é dicir, se a mostra escollida representa a totalidade da poboación (canto máis se aproxime o valor mostral ao valor real, máis calidade terá a estimación).

No caso de enquisas como a EPA, ou a maioría das elaboradas no INE, recórrase a procedementos complexos para a recollida da mostra e estimación dos parámetros o que afecta sustancialmente á varianza dos estimadores (Fay, R.E., e Train, G., 1995) Isto leva á aplicación de métodos máis elaborados para o cálculo de varianzas mostrais neste tipo de enquisas.

En España, para coñecer os erros da mostraxe na enquisa de poboación activa, emprégase o método usado para a estimación da varianza na “Current Population Survey” implantado pola Oficina do Censo de Estados Unidos.

Segundo nos din Azor Martínez, G., Jiménez Llorente, J.V., Pérez Arriero, C., Porras Puga, J. (2011), na EPA, o método empregado para obter os erros de mostraxe é o de “semimostras reiteradas” que consiste en obter sucesivas semimostras da mostra inicial.

A idea principal foi introducida por Mc Carthy en 1966 para o caso dunha mostra estratificada con dúas unidades de mostraxe por estrato (é o chamado “Balance Half Sample method- BHS), e basicamente pode describirse como o intento de capturar a variabilidade da mostra usando subconxuntos da propia mostra, para os que se replican os cálculos.

Neste caso o método consiste en que a partir de cada semimostra calcúlase a estimación da característica da que queremos obter o erro de mostraxe. Unha vez calculadas todas as estimacións con cada unha das submostras, así como a estimación coa mostra completa, o estimador da varianza vén dado pola seguinte expresión, segundo Azor Martínez, G., Jiménez Llorente, J.V., Pérez Arriero, C., Porras Puga, J. (2011) :

$$\hat{V}(\hat{\theta}) = \frac{1}{r} \sum_{i=1}^r (\hat{\theta}_i - \hat{\theta})^2$$

Sendo:

- r : o número de semimostras obtidas (o número de reiteracións).
- $\hat{\theta}_i$: a estimación obtida ca i -ésima reiteración
- $\hat{\theta}$: é a estimación baseada na mostra completa.

Para a nosa aplicación pártese dos datos da enquisa de poboación activa do terceiro trimestre de 2009, recollidos do IGE (2013). Dispoñemos dun total de 21291 datos (que é o tamaño mostral da EPA en Galicia), cos cales traballaremos para obter o erro de mostraxe.

Os datos están referidos a características específicas como as seguintes: provincia, tipo de rexistro das persoas da vivenda, datos demográficos (idade, relación de parentesco coa persoa de referencia, sexo, identificación do cónxuxe no fogar,

identificación do pai/nai no fogar, estado civil, lugar de nacemento, nivel de estudos, etc), datos referidos ao traballo (se traballou a cambio dunha remuneración na semana de referencia, traballou como “axuda familiar, datos do emprego principal, datos do segundo emprego, desexos de traballar máis horas, búsqueda de emprego, experiencia profesional, situación fai un ano, variables derivadas, factor de elevación, etc).

O obxectivo é medir o erro de mostraxe no relativo ó número de parados medido pola EPA, polo tanto as características interesantes, que van a permitir chegar a dito fin son as seguintes: a provincia (de cada unha das provincias faranse dúas submostras, ao ter catro provincias teremos oito submostras), o nivel (que nos indica se a persoa ten máis ou menos de 16 anos), o AOI (indícanos a relación coa actividade do entrevistado: se son ocupados por insuficiencia de horas ou se son o resto de ocupados, se son parados que buscan o primeiro emprego ou se xa traballaron antes ou, finalmente, se son inactivos); tamén se terá en conta o factor de elevación debido a que nos indica o número de elementos que hai na poboación por cada elemento da mostra.

Comezamos referíndonos ó total da mostra (a EPA en todo Galicia). Para a estimación da proporción de parados consideraremos como poboación obxectivo ó conxunto de persoas que forman parte do mercado de traballo. Polo tanto para esta análise o relevante será a poboación activa, polo que se eliminará da mostra os dous grupos de persoas que non entran neste concepto: os menores de 16 anos e os inactivos.

O seguinte paso será seleccionar os estratos cos que se vai traballar. No caso da EPA, os estratos selecciónanse segundo a cantidade de habitantes, pero os datos máis desagregados que dispoñemos son as provincias polo que serán os nosos estratos, sendo conscientes de que nos afastamos do cálculo que se fai na realidade. Traballamos logo con catro estratos, un por cada provincia.

A aplicación do método por provincias asimílase máis ao que fai EEUU, xa que o realiza por entidades xeográficas máis similares ás provincias (aínda que nos EEUU son máis desagregadas).

(Anexo 6: Datos poboacionais e mostrais provinciais e no total de Galicia)

Na EPA realízase unha mostraxe estratificada, e nela a varianza dos estimadores é igual a suma das variacións das estimacións dos totais en cada estrato. Segundo nos di Pérez, C. (2010), a varianza do estimador na mostraxe estratificada será:

$$V(\hat{\theta}) = \sum_{h=1}^L W_h^2 \cdot (1 - f_h) \cdot \frac{N_h}{N_h - 1} \cdot \frac{P_h Q_h}{n_h}$$

No ámbito que estamos estudando, os elementos que forman a fórmula significan o seguinte:

N_h : tamaño da provincia h

N : tamaño da poboación total

$$W_h = \frac{N_h}{N}$$

f_h : índice mostral = mostra na provincia h / tamaño da poboación na provincia h

P_h : tasa de paro na provincia h

Q_h : 1-tasa de paro na provincia h

n_h : tamaño mostral na provincia h

Calculando todos os membros que compoñen a fórmula da varianza para cada provincia, obtemos a varianza de cada provincia, e logo sumando as varianzas de todas as provincias, chegaremos a obtención da varianza mostral.

(Anexo 7: Obtención das varianzas provinciais e da varianza mostral)

Ó traballar cunha mostra estratificada, pódese empregar a idea de replicacións de media mostra para xerar un número de “pseudo-réplicas”. A idea básica é construír dúas mostras independentes dentro de cada estrato. Pola combinación das medias mostras dentro de cada estrato coas medias mostras nos outros estratos, en todas as formas posibles; así é posible xerar un gran número de diferentes submostras. Este concepto foi denominado “mostras interpenetrantes” por Deming en 1960. (Richardson, A.J., Ampt, Elizabeth S., Meyburg, Arnim H., 1995)

No noso cálculo, facendo dúas semimostras por provincia danos un total de 2^4 (16) réplicas como máximo, que será o número de réplicas que usaremos. No caso do INE realizan 40 réplicas, dado que a súa estrutura é máis complexa, o que da lugar a un número de semimostras posibles moi elevado.

Tendo en conta que a varianza en mostraxe estratificada combina o cálculo das varianzas por estratos, realizamos ese cálculo en cada semimuestra provincial, e posteriormente combinamos no cálculo da varianza de cada réplica, seleccionando as semimostras provinciais apropiadas.

(Anexo 8: Obtención das varianzas de cada submostra)

(Anexo 9: Obtención das reiteracións coas combinacións das varianzas de cada submostra)

Por último, aplicamos a fórmula da varianza do estimador no método de submostras reiteradas, explicada con anterioridade:

$$\hat{V}(\hat{\theta}) = \frac{1}{16} \sum_{i=1}^{16} (\hat{\theta}_i - \hat{\theta})^2$$

(Anexo 10: Obtención de cada reiteración e da varianza, desviación típica e coeficiente de variación total do estimador)

Conclusións

Como se viña dicindo ata o momento, a Economía precisa da Estatística, xa que para levar a cabo as súas funcións, necesita datos e poder tratalos estatisticamente para conseguir resultados cos que decidir e sacar conclusións.

A análise realizada no traballo corrobora esta idea nunha área relevante como a do mercado de traballo (relevancia que se incrementa na actualidade como consecuencia do número de parados tan elevado), na que permite obter indicadores que aproximen os niveis reais cuns niveis apropiados de calidade. En concreto, a Teoría de Mostras aplicada para a consecución da Enquisa de Poboación Activa permítenos, enquisando só a unha parte da poboación, coñecer as características de toda ela. Existen outras fontes das que podemos obter información do mercado laboral como poden ser os censos de poboación ou os rexistros administrativos, pero a calidade destes en canto aproximación aos valores reais do mercado de traballo non é tan boa como a conseguida pola Enquisa de Poboación Activa, a cal fai que o enquisado proporcione información máis veraz xa que non está sometido a posibles conveniencias de tipo económicas ou administrativas.

Debido a que estamos diante dunha estatística complexa, os métodos de mostraxe empregados son os máis elaborados. A maioría dos países analizados realizan un tipo de mostraxe similar ao efectuado en España, consistente nun estratificado en dúas etapas.

Nos países da Unión Europea, a Enquisa de Poboación Activa está normalizada en gran medida para todos eles polo EUROSTAT que se encarga de proporcionar definicións dos distintos tipos de actividade comúns para todos os membros. Isto implica que na maioría dos países da Unión Europea se realice o mesmo tipo de mostraxe.

Despois de analizar cada un dos posibles métodos a aplicar, concluímos que a mostraxe estratificada resulta a máis ventaxosa para obter resultados do mercado laboral; debido a que ao dividir á poboación en grupos en función da área xeográfica evitamos que algunha área estea menos representada que outras.

Estados Unidos, como gran potencial mundial que é, non nos defrauda nos indicadores que posúe do mercado de traballo, pero aínda que son máis numerosos que en España, prodúcese una situación parecida respecto ao noso país na metodoloxía empregada.

A “Current Population Survey” é a enquisa realizada nos Estados Unidos con máis parecidos coa española, emprega o método de semimostros reiteradas. Todo apunta a que esta enquisa é exemplar xa que no noso país adaptouse o método que empregaban alí para analizar os erros de mostraxe.

A “American Community Survey” estima a forza de traballo para un período de tempo ou para outro segundo a poboación que haxa no territorio a estudar, deste xeito, realiza estimacións a un, tres ou cinco anos dependendo da cantidade de poboación que abrangue a área a analizar. En España, a EPA realízase trimestralmente para tódolos estratos, sen discriminar por cantidade de poboación.

A “Survey of Income and Program Participation” ten a finalidade de examinar a dinámica do cambio no emprego. En España, a EPA tamén pretende coñecer esta dinámica ao facer que cada vivenda permaneza na mostra durante seis trimestres consecutivos.

Analizando as aplicacións que se foron implantando para realizar a enquisa acorde a evolución da tecnoloxía, comparando España con Canadá, por exemplo, observamos a tardía utilización en España do sistema de “Computer Assisted Interviewing”.

En canto ós turnos de rotación, algún países como Brasil seguen unha metodoloxía, na cal cada fogar seleccionado permanece catro meses consecutivos sendo investigado, oito meses sen ser investigado, e despois deste período, é investigado

novamente por catro meses máis e finalmente é excluído da mostra. En España, non se leva a cabo deste modo, as vivendas permanecen seis trimestres consecutivos sendo investigadas e logo saen da mostra, polo que son investigados durante 18 meses. Ca outra metodoloxía son investigados durante 16 meses, isto trae vantaxes e inconvenientes, polo que aínda que o cansancio nos enquisados é menor debido a que están fora da mostra durante oito meses, co método español conséguese información dun maior período de tempo no cal podemos observar posibles dinámicas no cambio de emprego.

A calidade que nos proporciona a EPA, foi xustificada no presente traballo coa aplicación práctica da obtención dos erros de mostraxe a través do método de semimostros reiteradas. Observamos como cada minucioso detalle a ter en conta (como por exemplo, a inclusión de tódalas características posibles da poboación ou a posesión dun método formal e eficaz xunto coa súa correcta aplicación) fai que a EPA conteña un reducido erro de mostraxe e polo tanto sexa un indicador excelente das características do mercado laboral. Ademais, os constantes esforzos e as previsións futuras de seguir mellorando a enquisa indican que a calidade da EPA seguirá en aumento.

Anexos

Anexo 1: Información que se ten en conta a hora de facer a EPA nos distintos países da UE, no ano 2011.

País	Participación	Tamaño da mostra / Índice de mostraxe global	Ratio de resposta	Unidade primaria de mostraxe	Unidade final de mostraxe	Tipo de mostraxe
España	Obrigatoria	69000 vivendas por trimestre / 0,5 %	83.5%	Áreas xeográficas	Vivendas	Estratificado en dúas etapas
Bélxica	Obrigatoria	14600 fogares por trimestre / 0.3%	67.4%	Seccións estatísticas	Fogares	Estratificado en dúas etapas
Bulgaria	Voluntaria	19504 fogares por trimestre / 0,65%	79.9%	Distritos enumerados polo censo	Fogares	Estratificado por conglomerados en dúas etapas
Rep. Checa	Voluntaria	33900 unidades de vivenda por trimestre / 0,6%	80.6%	Áreas censais	Vivendas	Estratificado en dúas etapas
Dinamarca	Voluntaria	40000 individuos por trimestre / 0,9%	51.5%	-	Persoas	Estratificado
Alemaña	Obrigatoria	86000 fogares por trimestre / 0,25%	97.9%	-	Grupos (distritos de mostraxe) de vivendas, fogares e persoas	Estratificado
Estonia	Obrigatoria	3200 fogares por trimestre / 0,7%	62.5%	Persoas	Fogares	Estratificado sistemático

Análise da metodoloxía estatística empregada na Enquisa de Poboación Activa

País	Participación	Tamaño da mostra / Índice de mostraxe global	Ratio de resposta	Unidade primaria de mostraxe	Unidade final de mostraxe	Tipo de mostraxe
Irlanda	Voluntaria	39000 fogares por trimestre / 3%	79.7%	Censo de bloques (de 75 vivendas cada bloque)	Grupos de fogares (un bloque contén 15 fogares)	Estratificado en dúas etapas
Grecia	Obrigatoria	32600 fogares por trimestre / 0,85%	82.0%	Bloques	Vivendas	Estratificado en dúas etapas
Francia	Obrigatoria	67500 vivendas por trimestre / 0,25%	84.0%	Áreas xeográficas	Grupos de unidades de vivendas	Estratificado en dúas etapas
Italia	Obrigatoria	77920 fogares por trimestre / 0,3%	88.9%	Municipios	Fogares	Estratificado en dúas etapas
Chipre	Obrigatoria	5300 unidades de vivenda por trimestre / 1,3%	96.3%	Áreas/pobos enumerados polo censo	Vivendas	Estratificado en dúas etapas
Letonia	Voluntaria	6032 fogares por trimestre / 0,68%	66.7%	Áreas enumeradas polo censo	Fogares	Estratificado en dúas etapas
Lituania	Voluntaria	7700 fogares por trimestre / 0,7 %	83.8%	-	Dirección das persoas seleccionadas	Estratificado
Luxemburgo	Voluntaria	7500 fogares por trimestre / 3,6%	32.7%	-	Fogares	Aleatorio simple
Hungría	Voluntaria	36200 vivendas por trimestre / 0,92 %	84.8%	Localidades	Vivendas	Estratificado en múltiples etapas
Malta	Obrigatoria	3200 fogares por trimestre / 2,25%	69.2%	-	Fogares	Aleatorio sistemático nunha etapa
Países Baixos	Voluntaria	50000 fogares por trimestre / 0,7%	78.2%	- Municipios (unidades primarias) - Direccións de correo (unidades secundarias)	Fogares	Estratificado en tres etapas
Austria	Obrigatoria	22500 vivendas por trimestre / 0,6%	92.7%	-	Vivendas	Estratificado simple
Polonia	Voluntaria	54704 unidades de vivenda / 0,4%	75.6%	Grupos censais (ciudades) / enumeración de distritos (rural)	Vivendas	Estratificado en dúas etapas
Portugal	Obrigatoria	22554 unidades de vivenda por trimestre/ 0,6%	82.4%	Áreas xeográficas	Vivendas	Conglomerados en dúas etapas

País	Participación	Tamaño da mostra / Índice de mostraxe global	Ratio de resposta	Unidade primaria de mostraxe	Unidade final de mostraxe	Tipo de mostraxe
Romanía	Voluntaria	28080 unidades de vivenda por trimestre / 0,39%	93.0%	Seccións censais	Grupos de (3) vivendas	Estratificado en dúas etapas
Eslovenia	Voluntaria	19750 individuos por trimestre / 1,0%	78.3%	-	Direccións	Estratificado sistemático
Eslovaquia	Obrigatoria	10250 vivendas por trimestre / 0,6%	93.1%	Censo de unidades administrativas	Vivendas	Estratificado en dúas etapas
Finlandia	Voluntaria	36000 persoas por trimestre / 0,9%	76.1%	-	Persoas	Estratificado sistemático
Suecia	Voluntaria	29500 persoas por mes / 1,2 %	74.6%	-	Persoas	Estratificado nunha etapa
Reino Unido	Voluntaria	86900 fogares por trimestre / 3,3%	61.9%	-	Direccións postais/números de teléfono/unidades de fogares	Estratificado nunha etapa (aleatorio simple no norte de Irlanda)
Islandia	Voluntaria	4120 persoas por trimestre / 1,9%	83.5%	-	Persoas	Aleatorio simple
Noruega	Obrigatoria	24000 persoas (12000 familias) por trimestre / 0,7%	83.4%	-	Fogares	Aleatorio simple
Suíza	Obrigatoria	41000 persoas por trimestre / 0,6%	84.9%	Fogares (mostra principal) / persoas (mostra adicional de estranxeiros)	Persoas	Estratificado en dúas etapas
Croacia	Voluntaria	5616 vivendas por trimestre / 0,39%	75.6%	Áreas censais	Vivendas	Estratificado en dúas etapas
Antiga Rep. Iugoslava de Macedonia	Voluntaria	5000 fogares por trimestre / 0,9 %	84.4%	Distritos enumerados	Fogares	Estratificado en dúas etapas
Turquía	Obrigatoria	42250 fogares por trimestre / 0,23%	87.0%	Bloques de fogares	Vivendas	Estratificado de dous clúster (con 8 submostras)

Fonte: elaboración propia a partir de datos da Organización Internacional do Traballo (2013), EUROSTAT (2012) e as Oficinas de Estatística dos respectivos países

Anexo 2: Situacións nas que resulta apropiada cada enquisa nos Estados Unidos:

Nivel Xeográfico	Estimacións transversais				Estimacións lonxitudinais
	Medición mensual do estado do emprego		Medición anual do estado do emprego		
	Estatus básico	Clasificados polas características socioeconómicas	Estatus básico	Clasificados polas características socioeconómicas	
Estados Unidos	CPS	CPS	CPS	CPS ou ACS	SIPP
Estado	CPS	CPS (estados seleccionados)	CPS	CPS ou ACS	SIPP (estados seleccionados)
Subestado Área de 65000 ou máis persoas	LAUS se está dispoñible	Non dispoñible	CPS ou LAUS se está dispoñible, do contrario ACS	CPS or ACS	Non dispoñible
Subestado Área de 20000 a 65000	LAUS se está dispoñible	Non dispoñible	LAUS se está dispoñible, do contrario ACS	ACS	Non dispoñible
Subestado Área de menos de 20,000	LAUS se está dispoñible	Non dispoñible	LAUS se está dispoñible, do contrario ACS	ACS	Non dispoñible

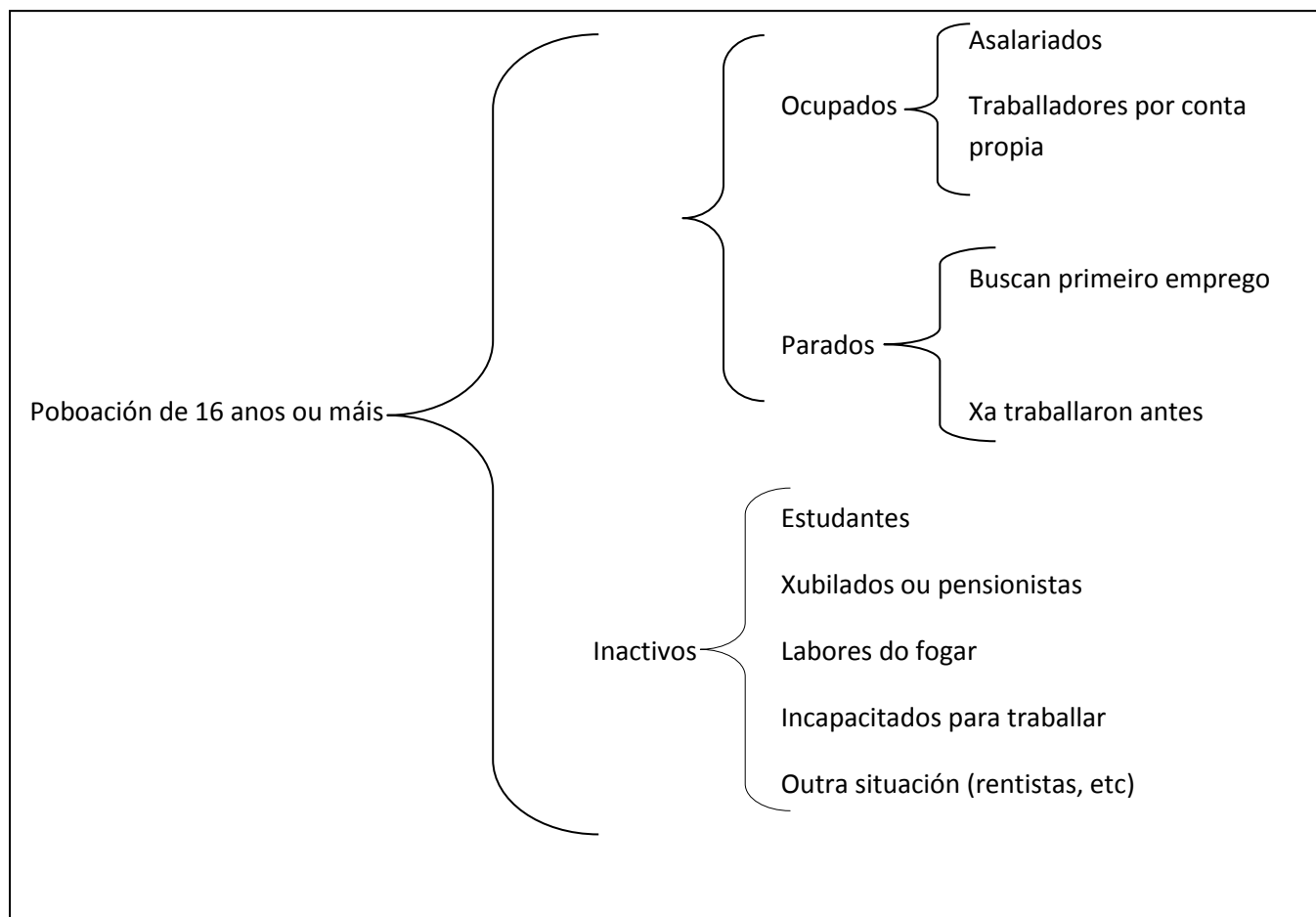
Fonte: elaboración propia a partir de conceptos recolectados na oficina do Censo de Estados Unidos.

Anexo 3: Porcentaxe de fogares en común, na mostra, en Brasil, entre pares de meses:

Intervalo de meses	Porcentaxe de fogares en común na mostra entre pares de meses
1	75,0
2	50,0
3	25,0
4 – 8	0,0
9	12,5
10	25,0
11	37,5
12	50,0
13	37,5
14	25,0
15	12,5
16 ou máis	0,0

Fonte: Instituto Brasileiro de Xeografía e Estatística

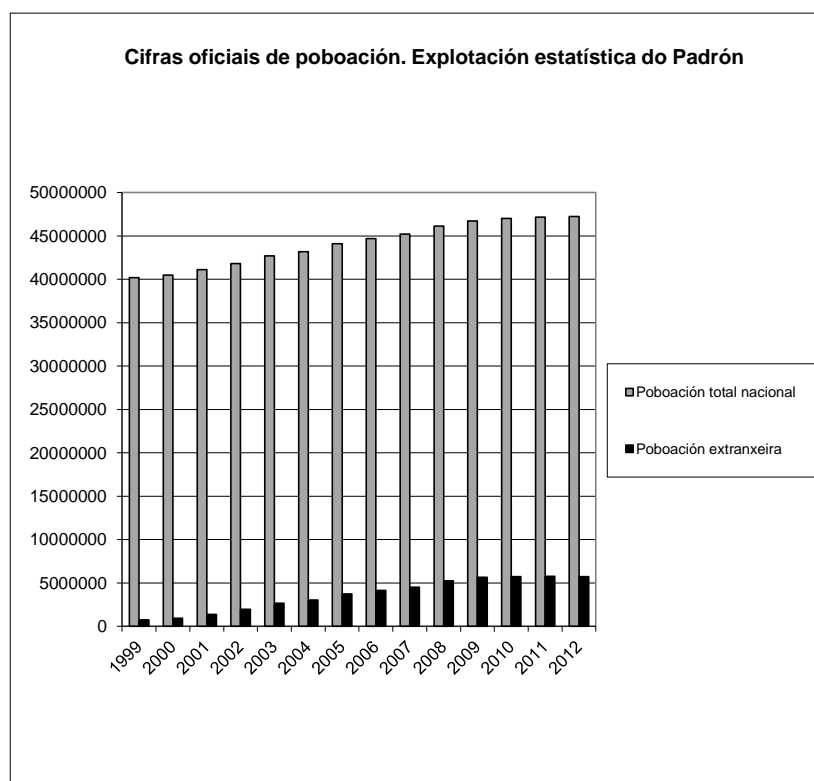
Anexo 4: Esquema de clasificación da poboación de 16 anos ou máis.



Fonte: elaboración propia.

Anexo 5: Aumento da poboación estranxeira a partir de 2001.

Ano	Poboación total nacional	Poboación estranxeira
1999	40202160	748953
2000	40499791	923879
2001	41116842	1370657
2002	41837894	1977946
2003	42717064	2664168
2004	43197684	3034326
2005	44108530	3730610
2006	44708964	4144166
2007	45200737	4519554
2008	46157822	5268762
2009	46745807	5648671
2010	47021031	5747734
2011	47190493	5751487
2012	47265321	5736258



Fonte: elaboración propia a partir de datos do INE

Anexo 6: Datos poboacionais e mostrais provinciais e no total de Galicia.

	A CORUÑA		LUGO		OURENSE		PONTEVEDRA	
Actividade	Fac. Elevación	Mostra	Fac. Elevación	Mostra	Fac. Elevación	Mostra	Fac. Elevación	Mostra
Ocupados subempregados	3329403	194	763066	66	732363	72	3487798	248
Resto de ocupados	46409163	2821	13459562	1340	10930592	1026	35997247	2748
Parados que buscan 1º emprego	540537	35	114579	12	146518	15	575130	46
Parados que xa traballaron antes	5795538	340	1156088	109	1365206	124	6525499	479
TOTAL	56074641	3390	15493295	1527	13174679	1237	46585674	3521

	GALICIA	
Actividade	FAC.ELEVACIÓN TOTAL	MOSTRA TOTAL
Ocupados subempregados	8312630	580
Resto de ocupados	106796564	7935
Parados que buscan 1º emprego	1376764	108
Parados que xa traballaron antes	14842331	1052
TOTAL	131328289	9675

Fonte: elaboración propia

Anexo 7: Obtención das varianzas provinciais e da varianza mostral

	A CORUÑA		LUGO		OURENSE		PONTEVEDRA	
	Poboación	Mostra	Poboación	Mostra	Poboación	Mostra	Poboación	Mostra
Ocupados	497386	3015	142226	1406	116630	1098	394850	2996
Parados	63361	375	12707	121	15117	139	71006	525
Total activos	560747	3390	154933	1527	131747	1237	465856	3521
Índice mostral	0,006045507		0,009855873		0,009389208		0,00755813	
PONDERACIÓN: Wh:Nh/N	0,426981085		0,117973811		0,100318819		0,354726285	
Taxa de paro (parados/activos)	0,112993917		0,082016097		0,114742651		0,15242049	
Varianza provincial	5,35755E-06		6,79466E-07		8,18646E-07		4,58195E-06	

	GALICIA	
	Poboación	Mostra
Ocupados	1151092	8515
Parados	162191	1160
Total	1313283	9675
Índice mostral	0,007367034	
Taxa de paro (parados/activos)	0,123500418	
Varianza total mostral	1,14376E-05	

Fonte: elaboración propia

Anexo 8: Obtención das varianzas de cada submostra

	A CORUÑA				LUGO			
	Submostra 1		Submostra 2		Submostra 1		Submostra 2	
	Poboación	Mostra	Poboación	Mostra	Poboación	Mostra	Poboación	Mostra
Ocupados	249682	1506	247704	1509	71179	705	71047	701
Parados	31890	189	31470	186	6059	58	6648	63
Total de activos	281572	1695	279174	1695	77238	763	77695	764
Índice mostrai	0,006019775		0,006071482		0,009878557		0,009833323	
Wh= Nh/N	4,253996072		4,217767034		1,166913431		1,173817797	
Taxa de paro	0,113257		0,112725397		0,078445843		0,085565352	
Varianza	0,001065778		0,001043355		0,000127743		0,000139724	

	OURENSE				PONTEVEDRA			
	Submostra 1		Submostra2		Submostra 1		Submostra 2	
	Poboación	Mostra	Poboación	Mostra	Poboación	Mostra	Poboación	Mostra
Ocupados	57448	541	59182	557	195572	1498	199279	1498
Parados	8742	78	6375	61	34986	262	36021	263
Total de activos	66190	619	65557	618	230558	1760	235300	1761
Índice mostrai	0,009351866		0,009426911		0,007633654		0,007484063	
Wh= Nh/N	1		0,990436622		3,483275419		3,554917661	
Taxa de paro	0,132074331		0,09724362		0,151744897		0,153085423	
Varianza	0,000183458		0,000138035		0,000880598		0,000923448	

Fonte: elaboración propia

Anexo 9: Obtención das reiteracións coas combinacións das varianzas de cada submostra.

	A CORUÑA		LUGO		OURENSE		PONTEVEDRA	
	Submostra 1	Submostra 2	Submostra 1	Submostra 2	Submostra 1	Submostra 2	Submostra 1	Submostra 2
Varianza	0,00106578	0,00104335	0,00012774	0,00013972	0,00018346	0,00013804	0,0008806	0,00092345
subm1	0,00106578	0	0,00012774	0	0,00018346	0	0,0008806	0
subm2	0,00106578	0	0,00012774	0	0,00018346	0	0	0,00092345
subm3	0,00106578	0	0,00012774	0	0	0,00013804	0,0008806	0
subm4	0,00106578	0	0,00012774	0	0	0,00013804	0	0,00092345
subm5	0,00106578	0	0	0,00013972	0,00018346	0	0,0008806	0
subm6	0,00106578	0	0	0,00013972	0,00018346	0	0	0,00092345
subm7	0,00106578	0	0	0,00013972	0	0,00013804	0,0008806	0
subm8	0,00106578	0	0	0,00013972	0	0,00013804	0	0,00092345
subm9	0	0,00104335	0,00012774	0	0,00018346	0	0,0008806	0
subm10	0	0,00104335	0,00012774	0	0,00018346	0	0	0,00092345
subm11	0	0,00104335	0,00012774	0	0	0,00013804	0,0008806	0
subm12	0	0,00104335	0,00012774	0	0	0,00013804	0	0,00092345
subm13	0	0,00104335	0	0,00013972	0,00018346	0	0,0008806	0
subm14	0	0,00104335	0	0,00013972	0,00018346	0	0	0,00092345
subm15	0	0,00104335	0	0,00013972	0	0,00013804	0,0008806	0
subm16	0	0,00104335	0	0,00013972	0	0,00013804	0	0,00092345

Fonte: elaboración propia

Anexo 10: Obtención da varianza de cada reiteración e da varianza, desviación típica e coeficiente de variación total do estimador.

Varianza de cada reiteración
0,002257576
0,002300427
0,002212153
0,002255004
0,002269557
0,002312408
0,002224135
0,002266985
0,002235153
0,002278004
0,002189731
0,002232581
0,002247135
0,002289985
0,002201712
0,002244563
Varianza total do estimador
5,01709E-06
Desviación típica do estimador
2,24E-03
Coeficiente de variación
0,018136

Fonte: elaboración propia

Bibliografía

- Alba Fernández, V. e Muñoz Vázquez, A. (2000). *Introducción a la estadística pública*. Jaén: Universidad de Jaén
- Alba Fernández, V. e Ruíz Fuentes, N. (2004). *Muestreo estadístico*. Oviedo: Septem Ediciones
- Azor Martínez, G., Jiménez Llorente, J.V., Pérez Arriero, C., Porras Puga, J. *Study of variance estimation methods in the Spanish Labour Force Survey (EPA)*. Recuperado de: http://www.ine.es/ss/Satellite?L=es_ES&c=INEDocTrabajo_C&cid=1259927978160&p=1254735116586&pagename=ProductosYServicios%2FPYSLayout
- Boletín Oficial del Estado (2008). *Convenio de colaboración entre el Instituto Nacional de Estadística (INE) y el Instituto Galego de Estatística (IGE)*. Recuperado de: http://www.boe.es/diario_boe/txt.php?id=BOE-A-2008-19504
- Casal, J. (2003). *Tipos de muestreo*. Barcelona: Universidad Autónoma de Barcelona.

- Cinco días (2013). *El desempleo de la EPA es el más real. Va a misa*. Recuperado o 8 de marzo de 2013 en http://cincodias.com/cincodias/2013/03/08/economia/1362761933_662691.html
- Fay, R.E., e Train, G. (1995). *Aspects of Survey and Model-Based Postcensal Estimation of Income and Poverty Characteristics for States and Counties*. Alexandria, VA: Proceedings of the Government Statistics Section. American Statistical Association, pp154-159.
- Hussmanns, R., Mehran, F., Verma, V. (1990). *Surveys of economically active population, employment, unemployment and underemployment: An ILO manual on concepts and methods*. Geneva: International Labour Office
- Instituto Galego de Estatística (2013). *Ficheiros de microdatos. Enquisa de poboación activa*. Recuperado de: http://www.ige.eu/web/mostrar_paxina.jsp?paxina=004002003002
- Instituto Nacional de Estadística (2005). *Cambios en la encuesta de población activa en 2005*. Recuperado de: <http://www.revistaindice.com/numero11/p6.pdf>
- Instituto Nacional de Estadística (2005). *Cifras INE: Boletín Informativa del Instituto Nacional de Estadística*. Recuperado de: <http://www.ine.es/revistas/cifraine/0105.pdf>
- Instituto Nacional de Estadística (2008). *Encuesta de Población Activa. Metodología 2005: Descripción de la encuesta, definiciones e instrucciones para la cumplimentación del cuestionario*. Recuperado de: <http://www.ine.es/daco/daco43/resumetepa.pdf>

- Instituto Nacional de Estadística (2012). *Encuesta de Población Activa: Diseño de la encuesta y evaluación de la calidad de los datos. Informe técnico*. Recuperado de: http://www.ine.es/docutrab/epa05_disenc/epa05_disenc.pdf
- Instituto Brasileiro de Geografia e Estatística (2001). *Mapa do Mercado de Trabalho no Brasil. Estudos e Pesquisas. Informação Demográfica e Socioeconômica*. Recuperado de: http://www.ibge.gov.br/home/estatistica/populacao/mapa_mercado_trabalho/mapa_mercado_trabalho.pdf
- Instituto Brasileiro de Geografia e Estatística (2007). *Série Relatórios Metodológicos. Pesquisa Mensal de Emprego. 2ª edição*. Recuperado de: [ftp://ftp.ibge.gov.br/Trabalho e Rendimento/Pesquisa Mensal de Emprego/Metodologia da Pesquisa/srmpme 2ed.pdf](ftp://ftp.ibge.gov.br/Trabalho_e_Rendimento/Pesquisa_Mensal_de_Emprego/Metodologia_da_Pesquisa/srmpme_2ed.pdf)
- International Labour Organization (2013). *Overview and topics of labour statistics*. Recuperado de: <http://www.ilo.org/global/statistics-and-databases/statistics-overview-and-topics/lang--en/index.htm>
- Levy, P. S., e Lemeshow, S. (1999). *Sampling of populations: Methods and applications*. (3rd ed.). New York: John Wiley and Sons.
- Martín Pliego, F.J. (1994). *Introducción a la Estadística Económica y Empresarial (Teoría y Práctica)*. Madrid: Editorial AC
- Mendenhall, W. e Reinmuth, J. (1978). *Estadística para administración y economía*. España: Grupo Editorial Iberoamérica
- Muruzábal Irigoyen, J.J. (2005). *Elementos de Estadística: Teoría de Muestras e Inferencia Estadística*. Madrid: Servicio de Publicaciones Colegio de Ingenieros de Caminos Canales y Puertos.

- Organización Internacional do Traballo (2013). *Sources and Methods: Labour Statistics*. Recuperado de: <http://laborsta.ilo.org/applv8/data/SSMe.html>
- Peña, Daniel (2001). *Fundamentos de estadística*. Madrid: Alianza Editorial
- Pérez, César (1999). *Técnicas de Muestreo Estadístico. Teoría, práctica y aplicaciones informáticas*. Madrid: ra-ma
- Pérez, César (2010). *Técnicas de Muestreo Estadístico*. Madrid: Garceta Grupo Editorial
- Richardson, A.J., Ampt, Elizabeth S., Meyburg, Arnim H. (1995). *Survey Methods for Transport Planning*. Eucalyptus Press. Recuperado de: [http://www.geog.ucsb.edu/~deutsch/geog111_211a/code_books/Survey Methods For Transport Planning.pdf](http://www.geog.ucsb.edu/~deutsch/geog111_211a/code_books/Survey_Methods_For_Transport_Planning.pdf)
- Rodríguez, R.J. (1996). *Teoría Básica del Muestreo*. Buenos Aires
- Sharon L.Lohr (1999). *Sampling: Design and Analysis*. Arizona State University: Duxbury Press
- Statistical Office of the European Communities (2012). *Labour force survey in the EU, candidate and EFTA countries. Main characteristics of national surveys, 2011*. doi:10.2785/35026
- Statistical Office of the European Communities (2013). *Employment and unemployment (LFS)*. Recuperado de: http://epp.eurostat.ec.europa.eu/portal/page/portal/employment_unemployment_lfs/methodology/definitions
- Statistics Canada (2013). *Labour Force Survey: definitions, data sources and methods*. Recuperado de:

http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3701&Item_Id=103716&lang=en

- Steven G. Medema e Warren J. Samuels. (1988). *A history of economic thought. Lionel Robbins*. United States of America: Princeton University Press and Oxford.
- Sue Zimmerman, Tamara (2012). Evaluation of National CPS Labor Force Variances. Washington, DC: Bureau of Labor Statistics. Recuperado de: <http://www.bls.gov/ore/pdf/st030090.pdf>
- United States Census Bureau (2013). *Guidance on Survey Differences in Labor Force Estimates*. Recuperado de: <http://www.census.gov/people/laborforce/publications/laborguidance092209.html>